

In Silico Structure-Based Prediction of Receptor–Ligand Binding Affinity: Current Progress and Challenges



Shailesh Kumar Panday and Indira Ghosh

Abstract Structure-based in silico studies aiming to predict affinity of a set of ligands to their cognate receptor have been enjoying keen interest and attention of researchers in drug design around the globe since many decades, and made significant progress to increase its predictive power, even it has emerged as a complementary field to in vivo and in vitro studies in recent years. Structure-based drug discovery (SBDD) process whose success heavily relies on a careful selection of structure of receptor and ligands and its accuracy, completeness, and rigor of chosen model, imitation of the physiological condition in such in silico models, e.g., pH and solvation. Appropriateness of selected mechanism of binding concept and the realization in mathematical terms used in scoring methods have a strong influence on the accuracy too. However, constant identification of new targets using systems approach like genomics, proteomics, metabolomics, and network biology has led a paradigm shift from single or a couple of targets toward the appreciation of emerging role of a network of targets. The application of such strategies in study of complex diseases is gaining attention. Identification of binding sites of receptor and their characterization is important to be able to portray its interacting features. It involves the search of ligands which are able to possess the features, present them complementary to the binding site, so by docking the set of ligands to the binding pocket of the receptor, activity can be evaluated. In silico receptor–ligand binding affinity prediction from docking has witnessed rigid-receptor rigid-ligand to flexible-ligand rigid-receptor treatment, and nowadays docking studies, through sampling side chain rotations of the binding site residues, also account for the flexibility of binding pocket of the receptor in indirect way. Literature survey has shown progress in ranking ligands in order of affinity using reliable scoring functions to find potent scaffolds which can be further optimized to gain more affinity. Many methods include effect of solvation in binding processes, like considering conserved water positions in active sites (water maps), explicit water simulation in presence of ligand with receptor, free energy perturbation, and thermodynamic

S. K. Panday · I. Ghosh (✉)

School of Computational and Integrative Sciences (SCIS),
Jawaharlal Nehru University, New Delhi 110067, India
e-mail: indira0654@gmail.com

© Springer Nature Switzerland AG 2019

C. G. Mohan (ed.), *Structural Bioinformatics: Applications in Preclinical Drug Discovery Process*, Challenges and Advances in Computational Chemistry and Physics 27, https://doi.org/10.1007/978-3-030-05282-9_5

109

integration. Availability of many conformers of receptors and ligands in solution suggests the importance of entropy in estimation of binding affinity, but entropy component of binding free energy directly is not included in such studies. In spite of unprecedented advancement of computational modeling, faster simulation techniques, accurate solvation models and current best practices, the dependence of binding affinity on pH, estimation of entropy along with enthalpy in binding affinity, inclusion of conformational entropy of ligand and receptor, and modulation of flexibilities during complex formation are important challenges lying ahead. Therefore, an account of prowess and challenges in structure-based prediction of binding affinity addressed in present review will provide directions for its appropriate application, understanding its limitations and getting important feedbacks for its betterment.

Keywords Structure-based drug design · X-ray crystal structure
Scoring function · Docking · Simulation · Structure validation
MM-PBSA · Entropy · Free energy

1 Introduction

The advancement of molecular understanding of the disease processes and their manifestations, along with computational advancement like *in silico* studies, aiming to predict high-affinity molecules/scaffolds binding to the target, grew as a promising complementary field of study mainly because of its cost-effectiveness and speed. It facilitated virtual high-throughput screening (vHTS) to narrow down the search space for further experimental work by making predictions about the ligand–receptor affinity [1]. Advancements in systems biology along with network biology helped identifying targets for diseases [2], and crystallography [3] and nuclear magnetic resonance (NMR) [4] techniques enabled solving structural models of the target molecules with higher resolution setting foundation of structure-based drug designing (SBDD). Docking is one of such computational studies, which aims to search high-affinity molecules from a library of chemicals and predict relative orientation (pose) of the molecule to the target. It also tries to rank the set of molecules/poses in a sorted affinity order [5]. Knowledge about the structure of receptors made binding site identification easier and enabled to screen the small-molecule libraries against the target seeking complementarity with the ligand.

Docking and scoring methods due to its promising applicability prospect has been extensively developed, critically evaluated, and constantly refined with the time, it has now shaped into a field of research; several software tools have been developed and are available for academic and industry research [5–11]. Recently, Taylor et al. [12] have reviewed the broad spectrum of major techniques amenable to the field of non-covalent docking studies, classifying them into molecular dynamics, Monte Carlo methods, genetic algorithms, fragment-based methods,

point complementarity methods, distance geometry methods, tabu searches, and systematic searches. They briefly presented algorithms and validations of models and techniques using test cases as examples. The study has concluded that hybrids of various types of algorithms employing novel search for appropriate poses and consensus scoring are better for large-scale docking [12]. It has been observed that rigid receptor and flexible ligand models achieved success rates of 70–80%. It can be influenced by the fact that programs implementing these algorithms were well established at that time [12]. However, they pointed out that possible reason for failure is underestimation of conformational sampling of receptor flexibility [12]. In spite of great success of docking methods in discriminating ligands as good and bad, predicting the binding on the basis of their affinity towards cognate receptor is poor. Moreover, in certain cases, docking shows inability to reproduce experimental binding pose and it is a great concern in the technical aspects of the docking methodology and its current progress, so need to review time to time. In 2010, Huang et al. [13] have discussed currently practiced docking techniques, delineating the ways for ligand sampling, accounting protein flexibility and specific scoring functions.

During a docking study, one has to do many sequences of tasks/steps which influence the final outcome of the study and its success [14]. First and the foremost thing is to search for the potential binding sites on the receptor and characterize them; however, sometimes when binding site is not known blind docking can be done. Several cavity detection algorithms and software were built to help this. In parallel, right selection of the receptor structure is crucial [14]; thus, the quality of the structure and experimental conditions used for resolving the structure has to be taken care of, and structure resolved with experimental conditions closest to the actual functioning condition should be preferred if available [15]. Most often, hydrogen atoms are missing in the structure; thus, protonation states of the titratable receptor residues have to be fixed, and usually, it is borrowed from predictions made using different protonation state prediction tools [16, 17]. Apart from the protonation states of titratable residues of the receptor, ionization states of ligands to be docked have influence on correct model of binding [16, 18]. Scoring functions also greatly influence the final outcome of the docking studies, and there are many scoring functions available; some may be suitable to study the specific type of protein active site but less effective in other cases [19]. Inherent demand of fast evaluation of poses during docking enforces the scoring functions to adopt approximations and parameterization, which compromises predictivity [19]. Thus, it is tough to guess which scoring will be suitable for which kind of active site. However, chemical intuition and consensus scoring protocols can be adopted to get better results.

Although the correctness of ranking and order of predicted affinity more often fail to provide significant correlation with experimental ranking and observed pose [20], such limitation of the *in silico* high-throughput screening can be partially attributed to the multifaceted problems in current practices, e.g., selection of appropriate binding theory, selection of appropriate modeling data, and limited knowledge about the reaction mechanism. Many such challenges are discussed in the present article.

1.1 Targets Are Diverse

To be able to comprehend the challenges lying ahead on the way to drug design/ drug discovery, it is important to understand the diversity of the drug targets that have been exploited so far as well as the trend in new drug targets in recent history of drug discovery [21]. Mathias Rask-Andersen et al. performed a study on all drugs approved by FDA during 1983–2010. They took all 1542 drug entries as on May 2009 and filtered out 225 drugs with unknown targets, 192 with no human targets, and 609 non-therapeutic targets to yield a dataset of 435 therapeutic effect-mediating targets for humans and to account for the time lag between drug approval and their entry in DrugBank; drugs approved during 2007–2010 were taken from FDA data and included for analysis. Drug–target association was annotated by manual curation from literature data, and targets were kept in four classes (receptors, enzymes, transporters, and others) with receptor class has highest 193 targets, followed by enzymes with 124, transporters with 67, and others with 51 targets [21]. Analyzing curated drug–target association dataset, they found that every year 17.9 drugs targeting human proteins are approved by FDA, while 4.3 of them act on novel targets. The trend in FDA approval of drugs targeting new human proteins (novel target drugs: NTDs) does not decrease overall. Moreover, they noticed three peaks corresponding to durations 1990–1993, 1994–2000, and 2001–2008 when NTDs were plotted against years from 1983 to 2010; they called them first-, second-, and third-target “innovation peaks,” respectively [21].

During the first innovation peak, it was observed that proportions of approved drugs for all major target groups—GPCRs, hydrolases, transferases, and isomerases—were similar to other two peaks. During second innovation peak, first time integrins appeared as drug target, while during the third innovation peak, asthma drug omalizumab-targeted Fc-receptors and imatinib appeared as kinase inhibitor [21].

Analysis of novel targets for drugs with time by Mathias Rask-Andersen et al. highlights the fact that with the passing time new drugs apart from targets belonging to earlier exploited classes, novel classes of targets are also being identified for new drugs. Thus, diversity in the classes of target molecules is expanding, and SBDD practices have to be optimized to improve success rates in such studies. Present review will attempt to enlighten and discuss the solutions for such relevant topics including the challenges upcoming ahead.

1.2 Targets Are More Diverse than Earlier

Genomic-wide association studies over a set of druggable genome, utilizing bioactivity data including approved drugs or clinical compounds and gene association data against these targets, can be used to come up with set of further druggable genes and gene combinations as target [22]. Recently in 2017, Finan et al. have

performed a similar study and estimated that 4479 genes can be drugged or are druggable out of total 20,300 annotated protein-coding genes as per Ensembl version 73 (<https://www.ensembl.org/>) covering ~22% of total. They reported that there could be 2282 genes more than earlier reports of the druggable human genome [22].

Systems biology approaches have been used for decades for predicting target genes in case of infectious diseases [2], studying systems approaches, e.g., metabolic control analysis (MCA) and flux balance analysis (FBA). Systems genetics approaches have also been used for identification of novel disease genes in rat and human [23]. Molecular networks information can be used for improving drug discovery projects at several stages from target identification utilizing information of existing data about drug–target association [24]. Metabolic and signaling pathway [25] and genome-wide association are studied in detail for identification of new target proteins and their interactions [26]. Genome-led methods provide a new pathway or a class of protein(s) as target.

Pharmacophore designed from ligands of a target protein can be looked for assessing binding site similarity for the proteins of same family as well as it can be used to compare binding site similarity for proteins from different families of proteins for selectivity. In recent times, several highly selective inhibitors of such protein(s) have been found to assess the multitarget activity. For example, c-Abl inhibitor imatinib [27] was approved as drug for chronic myeloid leukemia, but its clinical utility is widened after finding that it has shown significant activity against several other important targets, e.g., tyrosine-protein kinase kit (c-KIT or CD117). Similarly, sorafenib affects tumor proliferation and tumor angiogenesis pathways due to its multikinase inhibitory activity [28]. Sunitinib is also approved for being a multiprotein kinase inhibitor with similar effects as sorafenib [28].

1.3 Starting of Structure-Based Drug Design

One of the successful stories of the structure-based drug design started in the early eighties with purine nucleoside phosphorylase (PNP), targeted as a salvage enzyme important to inhibit, so that T-cell-mediated activation of immune system is suppressed. PNP is an important enzyme involved in purine salvage and catabolism [29]. Inactivity of PNP has been found to show adverse effect on T-cell proliferation [30]. Human PNP, a homotrimer with each subunit of molecular weight 97 kD, shows substrate specificity for guanine, inosine, and other 6-oxypurines analogs, while bacterial PNP shows specificity for adenine [30] also. PNP active site consists of three binding subsites: purine-binding site (Fig. 1, shown in cyan), hydrophobic site (or ribose-binding site, Fig. 1, shown in blue), and phosphate-binding site (Fig. 1, colored purple) [31]. In attempt to design potent PNP inhibitors, considering the features of three subsites of PNP binding site and three-dimensional structure of PNP as starting point, an iterative process of modeling inhibitor-bound structure, conformational search using Monte Carlo method followed by energy

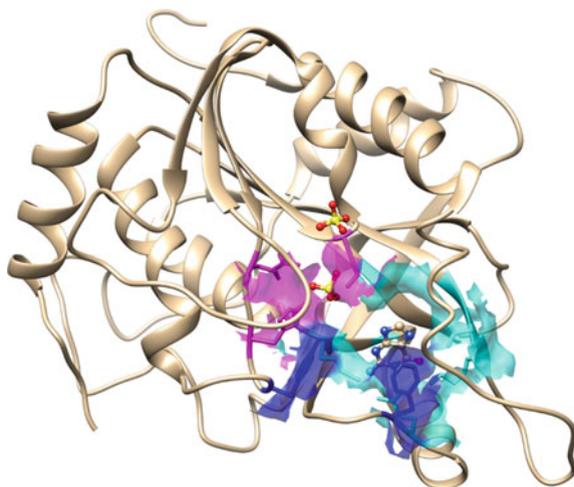


Fig. 1 Human purine nucleoside phosphorylase (PNP) monomer (PDB: 1ULB) in complex with guanine and sulfate ions. Guanine and sulfate ions are shown in ball and stick. Three subsites of PNP binding site: First subsite is called purine-binding site (shown in cyan surface, residues Ala116, Phe200, Glu201, Val217, Met219, Thr242, Asn243, Lys244), second subsite, i.e., hydrophobic site (or ribose-binding site consists of residues His86, Tyr88, Phe159 (from adjacent subunit of PNP trimer), Phe200, Met219) where Tyr88 and Phe200 are shown in blue surface. The third subsite termed phosphate-binding site (shown in purple surface residues Ser33, Arg84, His86, Ser220)

minimization and finally experimental determination of binding affinity and crystallization of complex structure was used. This iterative process yielded a series of potent and membrane-permeable 9-(arylmethyl)-9-deazapurines (2-amino-7-(arylmethyl)-4H-pyrrolo[3,2-d]-pyrimidin-4-ones) inhibitors of PNP [29]. Later, (S)-9-[1-(3-chlorophenyl)-2-carboxyethyl]-9-deazaguanin showed highest potency among all previously designed analogs [32]; however, the (R)-isomer was 30-fold less potent. This study exemplifies how structural information can be carefully used toward designing of potent inhibitors of the receptor of interest.

The enthalpy and entropy components of binding free energy together decide affinity of interaction between receptor and ligand. Therefore, affinity can be modulated favorably adopting following possible strategies: (i) decreasing the unfavorable entropy maintaining favorable enthalpy, (ii) increasing favorable enthalpy without introducing unfavorable entropy, and (iii) altering one or both of enthalpy and entropy favorably without losing proportionally on other component [33].

An example where first strategy has been used for optimizing affinity is inhibitors of PNP. Optimized picomolar-binding PNP inhibitors have also been reported [34]. The attention has been paid on reducing the entropic penalty, without sacrificing the enthalpy of binding to gain affinity. Hypoxanthine has K_i 4.3 μM , with enthalpy -30.5 kcal/mol, but 23.1 kcal/mol entropy penalty to result a -7.4 kcal/mol binding free energy [35], but optimized molecule SerMe-ImmH

shows 5.2 pM K_i , with -20.2 kcal/mol enthalpy, but merely 4.7 kcal/mol entropy to result -15.5 kcal/mol binding free energy [34].

The second strategy has been utilized for optimizing HIV-1 protease inhibitors. After the FDA approval of Indinavir in 1995, which binds only because of -14.2 kcal/mol entropy despite 1.8 kcal/mol unfavorable enthalpy with binding free energy -12.4 kcal/mol, the process of affinity optimization started. The constant optimization of inhibitors for efficacy leads to Darunavir which binds with only -2.3 kcal/mol favorable entropy; however, -12.7 kcal/mol favorable enthalpy yielded binding free energy -15.0 kcal/mol. The free energy gain of -2.6 kcal/mol was reported where every -1.4 kcal/mol results ten times better binder [36, 37]. Another such example involves cholesterol-lowering drug statins to HMG-CoA reductase, and Fluvastatin binds only due to -9.0 kcal/mol favorable entropy despite zero contribution from enthalpy. However, newer drug Rosuvastatin binding has only -3.0 kcal/mol entropy contributions, but additional -9.3 kcal/mol enthalpy gain results -12.3 kcal/mol binding free energy, -3.3 kcal/mol better than Fluvastatin [38].

The third strategy is more tedious and challenging mainly because of enthalpy entropy compensation, more often enthalpy can be increased by introducing new hydrogen bonding groups as a strong hydrogen bond which provides $\sim 4\text{--}5$ kcal/mol enthalpy; however, introduction of hydrogen bond decreases favorable solvation and entropy by structuring regions involved in hydrogen bonding. Alternatively, in theory, introducing multiple hydrogen bonds targeting same structural regions of receptor has been suggested to mitigate the extent of enthalpy entropy compensation [33].

1.4 Flexibility and Adaptability of Target

Initially, the protein–ligand docking was modeled as a lock-and-key, where protein was treated as “lock” containing a binding site as “key-hole” which can host a complementary ligand or “key.” However, later it was realized that lock-and-key model is not sufficient to characterize all binding events; thus, advanced models were proposed which can be put broadly in three groups: (i) lock-and-key (ii) induced fit (IF), and (iii) conformational selection (CS) [39]. The IF and CS models introduced to account for the receptor flexibility during the binding with ligands will be discussed in detail later. Although these models represent receptor–ligand binding in better way, still estimate only enthalpy of the interaction and the entropy component of the binding free energy remains to be estimated. It has been reported in the literature that entropic component of binding can be important in many interactions. A recent experimental and computational study of a human heat-shock protein 90 (HSP90) highlighted important alterations in binding properties of target on complex formation with small-molecule inhibitors [40]. Surprisingly, they found that compounds binding to helical conformation have increased target flexibility and gained entropy preference over compounds binding to loop conformation which was less flexible on complex formation [40].

1.5 Knowledge of Target Structure Is Essential but not Sufficient

In spite of success in structure-based drug discovery process [29] at several occasions, knowledge about the structure of the target involved in the disease does not necessarily lead to a drug for cure; β -Thalassemia is one such example. It is an inherited hematologic disease caused by less β -globin, largely reported in Mediterranean region, identified with the mutant β -globin [41]. The present treatment is continuous blood transfusions with chelation therapy [42] and less frequently, bone-marrow transplantation [43], because there is no drug treatment for cure. However, the first crystal structure of hemoglobin was known in 1968, and since then, more than 250 human hemoglobin structures are known [44]. Hence, druggability and understanding of disease is a field of research in itself, emerging as translational bioinformatics.

2 Challenges in Structure-Based Designing

As discussed in many review articles earlier, major steps to find in silico chemicals and design them for better inhibition of target macromolecule are identification of target protein or macromolecule of importance and associated functionally with the disease, characterization of its 3D structure and active site, mapping of interactions possible with chemical functional groups, docking, scoring, and finally ranking the possible chemicals to test experimentally. Each of these steps has many challenges which will be discussed here.

2.1 Accuracy of Structures

Before starting a docking study to screen, some library of compounds to come up with a set of molecules showing high binding affinity with the target receptor requires to have known 3D structure. The appropriate selection of the receptor structure can influence the success or failure of any screening study [14]. Therefore, a researcher needs a good structure to start with which could have been resolved mostly using X-ray or NMR. Sometimes, the structure of the desired receptor is not known. In such cases, a homology model of the structure can be used if a suitable template for the receptor can be found [14]. A template may be the same protein having similar function, showing high sequence similarity from different organism or even some other protein having same fold. If the structure of the receptor is known in advance, then there may be multiple structures resolved in different conditions, with varying resolution, varying model completeness, etc. In such a case, the most suitable structure has to be chosen [14]. In selecting receptor

structure, one has to keep in mind that how well the structure resolution condition matches with the actual functioning condition of the receptor and resolution of the structure [14]. Apart from this, many questions may arise like whether the structure is ligand bound? Whether active conformation of the structure is solved? Whether the structure is solved at pH similar to the functioning pH? These can also be of importance to consider during docking. The receptor crystal structure selection has to be done with care considering the quality of the structure model. Some of the most important parameters for crystal structure assessment have been outlined in the literature [45] and listed in Table 1. Crystal structure resolution which is a measure of quality of electron density data collected is one of such parameters; structures resolved at less than 1 Å are considered high-quality one being able to resolve electron densities at atomic level while structures greater than 3 Å have smeared electron densities and atomic positions are not clearly identifiable. Hence, crystal structure with resolution in range: 1 Å < resolution < 3 Å can be

Table 1 List of important parameters for assessing quality of X-ray crystal structure

Parameter	Description	Preferred	Comment
<i>Electron density and solved model quality</i>			
σ -cutoff	σ -cutoff applied to the data	None	
Lower resolution	A minimum spacing (d) of crystal lattice planes that still provide measurable diffraction of X-rays.	20–50 Å	
Higher resolution	A minimum spacing (d) of crystal lattice planes that still provide measurable diffraction of X-rays and also $\langle I/\sigma(I) \rangle$ greater than 2 in high-resolution shell.	<3 Å	Higher is better
Completeness	The number of observed reflections divided by the theoretical maximum	~100%	Higher is better
$\langle I/\sigma(I) \rangle$	The average ratio of reflection intensity to its estimated error. Signal-to-noise ratio	>2	
R-factor	A measure of the global reliability factor or goodness-of-fit between the experimentally obtained structure factor amplitudes, F_{obs} , and the calculated structure factor amplitudes, F_{calc} , obtained from the model.	<25%	Smaller is better
R_{free} -R-factor	R_{free} is R-factor for random ~5% reflections, not used for model refinement. R_{free} - R-factor < 2, may be indication of overfitting while R_{free} - R-factor > 7 may be due to poor refinement of model	2–7%	Smaller is better
R_{O2A}	Observation to atom ratio		Higher is better
<i>Geometric parameters of model quality</i>			
RMSD (bonds)	Root mean square deviation of bond lengths from ideal values	0.15–0.25 Å	
RMSD (angles)	Root mean square deviation of bond angles from ideal values	1°–3°	
Ramachandran violations	Number of ϕ - ψ torsion pairs falling in disallowed regions of Ramachandran plot	0	Smaller is better

considered reasonable quality structures [46]. Apart from resolution, R_{value} , R_{free} and real-space R-value and real-value correlations are among the important parameters to assess the quality of crystal structure as discussed by Brown et al., in 2007 [45].

Geometric parameters and quality of structure: Apart from diffraction quality and structure refinement parameters, geometric and chemical parameters are equally important to consider while assessing its quality [15]. Atomic positions in model, planarity of peptide plane, stereoisomer of peptide bond, bond length, bond angle, and torsions angles should be checked for an unnatural occurrence [15]. Since all combinations of backbone torsions ϕ - ψ cannot occur in proteins, only those pairs which conform to the Ramachandran plot, thus number of ϕ - ψ pairs in disallowed regions of the Ramachandran plot which ideally should be zero, generally lesser violation considered better structure, are used as a critical parameter for the quality of the crystal/model structure as best practices.

Atomic occupancy and B-factor are among other important parameters to be considered while assessing the quality of structure. Occupancy of an atom is the fraction of molecules which occupy modeled position among all molecules in crystal. An occupancy 0.0 means modeled positions not observed in crystal, and 1.0 means modeled position is present in all molecules in crystal [47]. If some residues in crystal structure show more than one conformations in crystal structure, then conformation with highest occupancy should be preferred. In case of ligands, the occupancy is dependent on K_d value, e.g., for a ligand with K_d in range 10–100 mM, maximum achievable occupancy ranges 70–90% or 0.70–0.90 considering working ligand concentration <500 mM [48]. B-factor in theory represents the amplitude of oscillation of the atom around equilibrium position. It quantifies the dynamics of the atom; often, isotropic B-factors are reported in crystal structures; however, anisotropic B-factors may be reported in high-resolution structures. For high-resolution structures, anisotropic atomic displacement parameter (B-factor) can be substantiated only when resolution is higher than $\sim 1.4 \text{ \AA}$ [46]. Structural regions in crystal structure having B-factor higher than a threshold B_{max} should be carefully inspected because of their implications to high disorder in the region [49].

At times, in crystal structure water molecules play important role in binding and have to be considered for characterizing the binding site for its water interaction profiles [50]. However, identification of structurally important waters involved in receptor–ligand interaction is another challenge [51, 52].

Proteins are usually flexible molecules, and inherent dynamics characterizes its interaction. Moreover, a crystal structure is usually a time and space average of the conformers present in the crystal lattice [15]. Therefore, quite often it may not be the conformation presenting the best possible affinity for the given ligand due to the rigid treatment of the receptor. Thus, protein should be allowed to flex in such way that it could show best possible affinity with the ligand.

2.2 Comparative Homology Modeling and Role of Template

Very often the target protein crystallization is not possible, and no other way but homology or comparative modeling of structure becomes imperative. Many standard tools and directions are reviewed, and appropriate protocols are included [53, 54]. Many such tools to evaluate the modeled structures are also discussed in the literature [55, 56]. Here we shall cite a specific example showing importance of choice of template using homology modeling applied for *Mtb* isocitrate dehydrogenase (ICD).

Mycobacterium tuberculosis is known to use the glyoxylate shunt during the persistent stage [57]. Experiments have been performed to understand the glyoxylate shunt by considering the close analogy with *Escherichia coli* system [58]. For *E. coli*, glyoxylate shunt pathway is well studied and is initiated by phosphorylation of specific serine-105 residue of isocitrate dehydrogenase (ICD) [59]. *Mycobacterium tuberculosis* being a prokaryotic organism, same type of functionality was also expected for the glyoxylate bypass pathway [58, 60].

Phylogenetic analysis of the ICD sequences shows that *Mtb* has NADP-dependent ICD which belongs to subfamily II of ICD. Subfamily II has predominantly eukaryotic members, while *E. coli* ICD is classified in subfamily I [61]. Across the family, ICDs are found to be functional either monomers or dimers. *E. coli*, *Mtb*, and human all have functional homodimeric forms. Dimeric ICDs contain active sites which are contributed by the residues of both domains. Though *Mtb* ICD is regulated by phosphorylation process, it is more equivalent to eukaryotic ICDs. Eukaryotic ICDs are not found to be regulated by the phosphorylation, and also mammalian system does not possess glyoxylate shunt [62]. So overall evidence suggest that *Mtb* ICD has close similarity with eukaryotic system; however, the presence of glyoxylate shunt pathway makes this system closer to prokaryotic intracellular pathogenic survivor.

Understanding of shunt pathway shown that regulation of the *Mtb*'s ICD depends upon the phosphorylation/de-phosphorylation state which is expected to be regulated by some of available 11 serine/threonine phosphatase/kinases [63]. In 2009, Vinekar et al. had performed molecular dynamics simulation-based analysis to understand the effect of selective phosphorylation of serine residues [62]. However, crystal structure of *Mtb* ICD was not available at that time (Table 2), so homology modeling had been done using different crystal structures as templates to select appropriate functional model.

The ultimate goal of the homology-based structure modeling is to model the structure from its sequence with an accuracy that is comparable to the best results achieved experimentally. As the crystal structure of *Mtb* ICD was unavailable, homology-based structure modeling was the preferred way to understand the structural features of the ICD. For ICD modeling, target sequence (UniProt ID: P9WKL1) was found to align with many sequences of already crystallized structures from both prokaryote and eukaryote. Based on the homology rules of %-identity, functionality, quality of the structure, and association with same taxonomy, three ICDs [64] were

Table 2 Comparison of the crystal structure of *Mtb*^a with selected (template) prokaryote and eukaryote crystal structures

ICDH	<i>Mtb</i>	<i>Sus scrofa</i>	<i>E. coli</i>
Sequence length	409	413	416
Template PDB ID	4HCX [66]	1LWD [65]	3ICD [64]
Year of publication	2013	2002	1989
Template structure resolution (Å)	2.18	1.85	2.5
R_{free}	0.262	1.85 Å	NA
R_{work}	0.205	0.210	0.180
Ramachandran outliers (%)	1.8	0.2	0.5
Sequence Identity with respect to <i>Mtb</i> ICDH (UniProt ID: P9WKL1) (%)	100	65.2	23.6
Sequence Similarity with respect to <i>Mtb</i> ICDH (UniProt ID: P9WKL1) (%)	100	79.2	35.7

^aEarlier modeled because structure was not available till 2013

selected as template structure for modeling. However, in cross-taxonomy (with eukaryote) 1LWD [65], same target sequence had higher sequence identity (Table 2) than *E. coli*. Both crystal structures (3ICD and 1LWD) have same Rossmann fold and a common dinucleotide-binding domain [64, 65].

In such case, where target structure from the same taxonomy is available and fulfills the most homology modeling criteria, it is not always true that model structure will also provide functional explanation. Model developed using *E. coli* is shown in Fig. 2a (dark gray color) with *E. coli* crystal structure (green color). Both structures are superimposed well with RMSD 4.68 Å. However, model structure (white color) developed using *Sus scrofa* (orange color) as template superimposes with crystal structure with RMSD of 0.57 Å (Fig. 2c). Both models are validated using PROCHECK [55], and more than 85% residues are found under Ramachandran region. So, both models follow homology criterion and passed by the structure validation tools.

In 2013, Quartararo et al. published the crystal structure of *Mtb* ICD dimer complex with NADPH. This structure is then used to understand the closeness of modeled structure of *Mtb* ICD with both *E. coli* and *Sus scrofa*. Superimposition of *Mtb* with *E. coli* and *Sus scrofa* is shown in Fig. 2b, d, respectively. Although all three have same folds, *Sus scrofa* is more close toward *Mtb* than *E. coli*. *E. coli* structure has 6.4 Å RMSD with *Mtb*, and major differences occurred in the beta-hairpin loop region where *E. coli* structure has helical element than beta-structure element. This region of dissimilarity known as clasp region between inter-subunit interface [64] plays important functional role during phosphorylation [61].

So, from this case study, it is very clear that one template cannot guarantee about the functional state of the homology model, so different templates may be used to develop appropriate functional model, as mentioned in comparative modeling review [53]. Key to the selection of the model is always to be associated with the

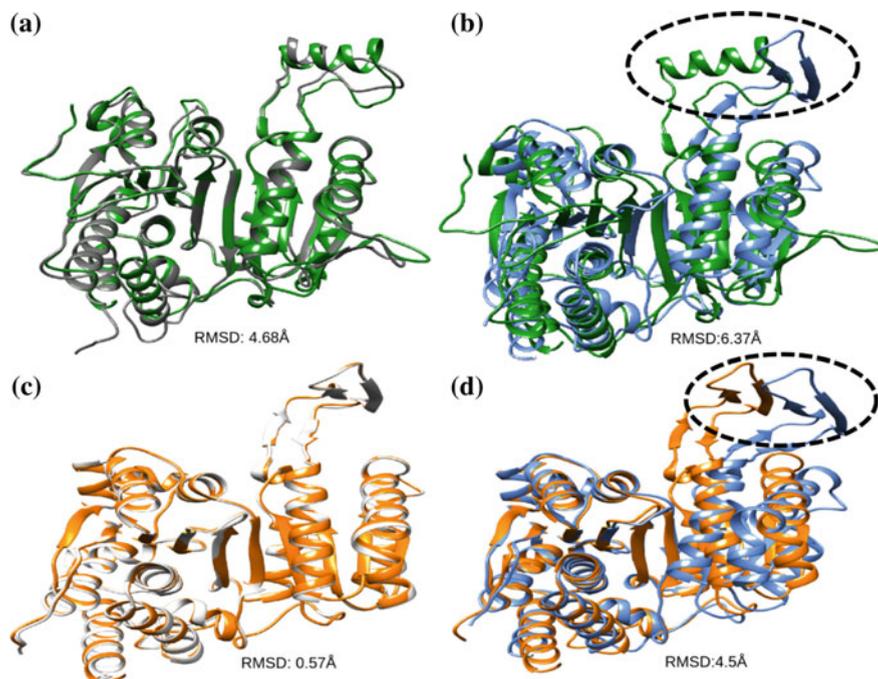


Fig. 2 Two homology-based models have been developed for *Mtb* ICD using two different crystal structures (one from *E. coli* and one from *Sus scrofa*). **a** Shows the modeled structure (dark gray color) superimposed with *E. coli* crystal structure (green color). Model fit well with 4.6 Å RMSD value. **c** Second model is developed using *Sus scrofa* structure (1LWD) and superimposed model structure (white color) is shown with 1LWD (orange color). When *Mtb* structure published in 2013, it is found that mammalian ICD is much closure to *Mtb* as shown in panel **(d)** than *E. coli* (panel **b**). Fold is well conserved in both models, but major differences are highlighted in clasp region (shown in black circle)

experimentally known functional features. It is also established that structure validation tools like PROCHECK [55] and WHAT IF [56] can only suggest the quality of the models not the functionality of the modeled protein. Other methods popularly known as ab initio designing of protein, alternate to template-based modeling, have been discussed in other reviews [67, 68]. A comparison of efficiency of modeling protein structure called CASP (critical assessment of methods of protein structure prediction) provides evaluation of such programs [69]. Recently, designing of protein structures has been successfully applied to model protein from genome sequence using an integrated pipeline by Jayaram and co-workers [70]. However, ensembles of model structure may provide a better docking success which has been cited in 2010 by Novoa et al. [71].

2.3 Ligand Flexibility

Apart from the traditional approach to look for potential inhibitor as small molecules for proteins, small peptides can also be strategically designed to complement interaction hot spots presented by receptor molecules, using knowledge about the structure of receptor and its interacting partner molecules. In a recent article published in *Science*, Kadam et al. [72] have exemplified the approach. The study focuses on influenza type 1 virus and their surface protein hemagglutinin (HA), which is associated with virus invasion of host cells. HA is composed of two domains HA1 and HA2, and functional unit is a homotrimer of HA. The interface of HA1 and HA2 forms a hydrophobic pocket. This HA-binding site, which is near the stem region of the HA membrane, is targeted by the broadly neutralizing antibodies (bnAbs) of the host and blocks large conformational rearrangement associated with membrane fusion and thus neutralize virus [72]. Structurally, analyzing the epitopes, at HA1/HA2 interface, a highly conserved site was found. This structural information allowed researchers to synthesize novel proteins, e.g., HB80 and HB36, which could mimic bnAb paratope CR6261 and bind in the conserved hydrophobic pocket, by placing amino acid side chains in appropriate configuration and conformation. These proteins did show binding affinity comparable to CR6261 and inhibited low pH-induced conformational change in HA. Further, optimizations lead to improved analogues of HB36, which were also effective in protecting mice against lethal H1N1 infections [72].

Success of de novo designed protein inspired researcher to look for even smaller peptide like inhibitors seeking better drug-like properties, e.g., availability in blood stream with higher lifetime. Starting from the available structural and functional information about bnAbs, e.g., CR9114, CR6261, F10, A06, FI6v3, HCDR3 was selected which possesses major interactions as the starting point for design of smaller HA inhibitory peptides. After creating a pool of potential HA inhibitory peptides mimicking different structural features of the HCDR3 loop [72] and characterization of each peptide in terms of its thermodynamic (K_d) and kinetic parameters (k_{off} and $t_{1/2}$), a combination of all distinct structural features of these peptides into an 11-mer cyclic peptide containing five non-proteinogenic residues was synthesized. This peptide showed better affinity and longer residence time for binding to HA. This study exemplified a novel approach, where compendium of available structure is utilized with chemical intuition of structure and function to yield a small cyclic peptide with better therapeutic prospect over existing inhibitory proteins, e.g., HB36 and its variants [72].

Alternatively, another novel idea has been floated by Young et al. of stapling small peptides to protect them from proteolytic cleavage and further designed a series of stapled peptides among which mimic of α -helical peptide ATSP-7041 was reported to be a potent and selective dual inhibitor of MDMX and MDM2 [73]. However, MDM2 and MDMX are suppressor of p53, thereby activates p53 pathway in tumors [74]. In a recent in silico study, where Garima et al. tried to study the mechanistic aspect of recognition of small stapled α -helical peptide ATSP-7041 with human serum albumin

(HSA) and compared it with mouse serum albumin (MSA) [75], starting from the crystal structures of HSA and ATSP-7041 in complex with MDMX. They used 50 ns molecular dynamics simulations to sample conformational states of HSA; simulation trajectories were clustered to give five clusters, and in these six (five cluster representatives and one crystal structure) HSA conformations were used for further docking studies. ATSP-7041 were fully blindly docked to above six HSA conformations using protein–peptide docking tool pepATTRACT [76] and generated ensemble ($\sim 24,000$ poses) of possible docking poses for each; then these ensemble of poses was clustered using k -means algorithm to result 40 clusters for each of six HSA conformations. Further, they refined each of the 40 clusters representative poses for each of six HSA conformations and then performed MD simulation for 5 ns to assess the stability of the pose. Their study resulted four binding sites R1, R2, R3, and R4 which were most occupied and considered for further study. Moreover, representative poses of ATSP-7041 and HSA complex one for each site was simulated using explicit solvent, and binding affinity was estimated using MM-GBSA method. However, for MSA, no crystal structure was available, so they modeled it using swiss model choosing HSA as template. ATSP-7041 was kept in MSA at sites R1, R2, R3, and R4, and three replicates of 100 ns MD simulation in explicit solvent were performed. Their analysis of these results suggested that sites R2 and R3 were not stable for mouse in contrast to human which they attributed to sequence dissimilarity at the region in human and mouse serum albumins. Moreover, they also found that sites R1 and R4 have lesser affinity in case of mouse for ATSP-7041 serum albumin binding than HSA. They also predicted a list of residues in the binding pocket contribution to the difference in binding energy. The binding site R1 is canonical binding site overlaps with already known site called Sudlow's site II, but R4 appears to be a novel binding site. Such in silico studies try to provide computational protocols which can be carefully utilized to gain mechanistic detail into protein–ligand interaction processes. Flexible ligands, e.g., peptides, can show better complementarity by conformational adaptation to attain several weak interactions with the receptor [77]. Potential to gain affinity through modulation of flexibility of ligands has been sensed, and nowadays, smaller peptides are also being evaluated by researchers across the globe for their therapeutic usage.

2.4 Protein Flexibility During Binding

Proteins are generally flexible molecules. Therefore, flexibility of the receptor has to be accounted in in silico binding affinity prediction studies to better represent the physicochemical conditions. The enormous conformational space available to proteins is very challenging to exhaust in docking studies because of unrealistic sampling requirements. However, non-exhaustive but simplistic and computationally less demanding methods have been developed over the years as proxy for accounting the flexibility of the protein during the binding which can broadly be put in four classes: soft docking, side chain rotation, molecular relaxation, and docking to multiple structures.

Soft Docking: This technique allows small conformational relaxations by treating van der Waals which overlaps through a softened potential and is efficient in terms of computational cost, but it can only account for smaller relaxation in receptor structure during binding to ligand [78]. Ferrari et al. [78] applied this method using two cavities of T4 lysozyme and drug-target aldose reductase which undergo large conformation change during binding. Available Chemicals Directory (ACD) [78] was screened against chosen targets for evaluating the method. They reported, with single receptor conformation, soft potential was better in identifying known ligands, while with multiple receptor conformations, it was poor in identifying leads than hard function; this trend was similar for both receptor and more pronounced for aldose reductase. Soft docking gives better score for ligands and decoys thereby better scoring, but it misses true ligands [78]. Qualitatively, similar results were reported by soft-docking studies of protein–protein [79] and antigen–antibody [80] interaction studies.

Side chain rotation: Allowing side chains rotation of the binding site residues of the receptor is computationally costlier than soft docking but offers better ways to account flexibility of receptor through sampling side chain rotations of binding site residues and overcome the limitations of soft docking, avoiding unphysical van der Waals clashes in predicted poses [81]. Preliminary idea of incorporating side chain flexibility into docking through usage of rotamer states of the binding site residues with rigid ligand conformation by Leach et al. [82] has been carried forward and adapted in several studies. For example, approach of rigid anchor and flexible complementary growth of ligand in receptor-binding site is implemented in SLIDE by Schnecke et al. [83] and used it to screen for potential ligands of progesterone receptor, dihydrofolate reductase, and a DNA-repair enzyme from a dataset of 175,000 organic compounds. Another approach introduced by Dean and co-workers [84] is applied to successfully reproduce experimental pose of ligand in binding site by docking synthetic inhibitor RS-104966 to the S1' pocket of the human collagenase matrix metalloproteinase 1 (MMP-1) [84]. In this approach, an ensemble of binding site conformations was generated using side chain rotamer states of the binding site residues followed by identification of representative conformations combining principal component analysis and fuzzy clustering [84]. Frimurer et al. performed a study attempting to assess the extent of impact of flexible side chain conformations of binding site residues on predicted binding poses and affinity [85]. They chose protein, phosphatase tyrosine 1B co-crystalized with non-peptide inhibitors, and docked ligands to parent receptor structure, resulting correct poses to correlate with low predicted binding energy [85]. In the process, an ensemble of structures was generated using rotameric states of subset of binding site residues (Asp48, Lys120, and Phe182), and ligands were docked to each structure; correlation of binding affinity with predicted scores improved for correct poses [85]. The importance of considering side chain flexibility in docking is also highlighted in study of Gaudreault et al. They created a curated non-redundant dataset of 188 proteins where unbound- and bound-both structures were already crystallized. In their study, they found that 90% binding sites and side chain rotation were accounting the flexibility in it, and 30% of them were essential side chain rotation and only 10% binding sites are rigid [86].

Molecular relaxation: This concept takes one step further toward accounting protein flexibility from side chain rotation. In this approach, ligand is docked in the binding site of the receptor allowing potential atomic overlaps to certain extent followed by relaxation stage where docked pose of the ligand is energy minimized and complex is relaxed allowing backbone relaxation along with side chain using molecular dynamics or Monte Carlo simulation. Apostolakis et al. performed a study in which they tried to incorporate receptor flexibility to model induced fit in ligand and binding site over three challenging docking cases: (i) anti-steroid antibody DB3 with two ligands, a rigid-ligand progesterone (no rotatable bonds) and (ii) a flexible-ligand 5β -androstane-3,17-dione (having rotatable bonds), and (iii) N^{α} -(2-naphthyl-sulfonyl-glycyl)-D-*para*-amidino-phenyl-alanyl-piperidine (NAPAP) binding to human α -thrombin [87]. Progesterone and 5β -androstane-3,17-dione show two different binding modes, thus make a perfect test case. In this method, ligand was seeded to the center of binding pocket in random pose followed by a combination of minimization with shifted non-bonded interaction and Monte Carlo minimization; authors were able to successfully reproduce the crystalized pose for test cases with native structure of protein and without prior knowledge of structure of NAPAP in α -thrombin case [87]. This study highlighted the importance of considering receptor flexibility under the influence of ligands interaction field in docking. Davis and Baker [88] implemented a method in ROSETTALIGAND to account for the receptor backbone flexibility along with full-ligand flexibility and showed that on a challenging cross-docking test case of Meiler and Baker [89] (10 co-crystallized receptor–ligand pairs, with large flexible ligands and multiple side chains with changing rotamer), their new method reproduces binding poses better (lower RMSD for best-scoring docked poses) in comparison to their rigid-backbone docking.

Multiple structure docking: McCammon and co-workers [90] used relaxed complex method to dock fully flexible version of prospective drug molecules JE-2147 wild-type and V82F/I84V drug-resistant mutants of HIV-1 protease ensemble of conformations. In both cases, wild-type and mutant HIV-1 protease, an ensemble of 2200 conformation from 22 ns all atom explicit solvent MD simulation of closed conformers of apo structures of receptor and coordinates were saved every 10 ps; in both cases, crystal structure poses were successfully reproduced. Later, JE-2147 was docked to each 2200 conformation for both wild-type and mutant cases and optimized the protocol. To synthesize test inhibitors, same protocol was applied to dock 23 newly designed potential inhibitor (called JE.D.I. series molecules) to 700 conformations of the HIV-1 protease mutant. Based on high binding free energy of four compounds of the JE.D.I., which were significantly different from their parent compound JE-2147 as well rest members of the series; four new compounds with potentially better pharmacological properties were suggested for test [90].

Similar concept but using MD simulation to dock and identify the interactions between domain motions to influence the inhibitor/ligand binding has been attempted in case of Fe-artemisinin adduct binding to PfATP6, a Ca^{++} transporter well-known target in *Plasmodium falciparum* [91].

Sarco-endoplasmic reticulum membrane calcium ATPase (SERCA) is Ca^{++} transporting ATPase; it is found in the mammalian systems and regulate the Ca^{++} flow between cytoplasm and membrane-bound stores [92]. SERCA-type transporter is also found in *P. falciparum* and is known as PfATP6. PfATP6 is large multidomain Ca^{++} channel receptor and only orthologous receptor to mammalian SERCA [92]. Importance of this channel receptor highlighted in 2003 when it was found that artemisinin (one of the most effective antimalarial drug) targets this receptor [93]. To understand the plausible mechanism of artemisinin action on PfATP6, extensive molecular dynamics simulation-based study has been performed [91]. This computational study shows that activated artemisinin (Fe-Artemisinin adduct) enforced large conformational changes in the extracellular domains (Fig. 3). Artemisinin adduct binds in the membrane-bound helical region and makes a hydrogen bond network which connects it with extracellular nucleotide

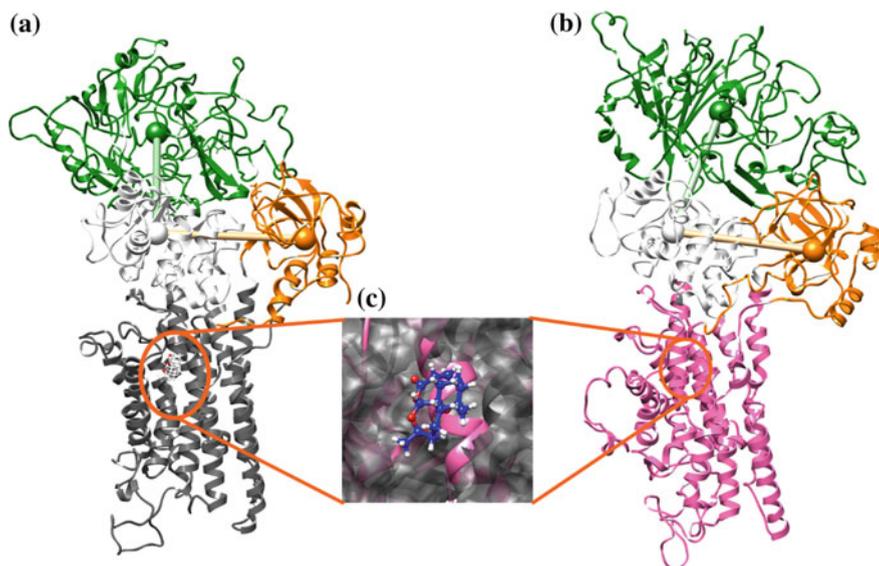


Fig. 3 Importance of receptor flexibility as observed in case of Fe-artemisinin adduct binding to *Plasmodium falciparum* ATP6 (PfATP6). Region spanning residues 364–799 shown in green contains nucleotide domain (N), region of residues 1–45 and 130–253 shown in orange contains actuator domain (A), region of residues 800 to 959 shown in white contains phosphorylation domain (P), and transmembrane region is shown in dark gray and pink colors in panel A and B, respectively. Ca^{++} and ligand binding sites are in the transmembrane region. Centroids of domain N, P, and A domains are shown with green, white, and orange spheres, respectively. The angle between centroid of domains N-P-A comes down to 78.5° (panel B) from 89.6° in open form (panel A), and distance N-A in closed conformation comes down to 44.9\AA from open conformation distance 53.7\AA (see panels B and A, respectively). **a** Open-form receptor is shown in ribbon, Fe-artemisinin adduct in ball and stick with carbons in white and rest atoms colored by atom types. **b** Shows closed form or receptor; **c** dark gray surface shows ligand-binding site in open form, and pink ribbon shows closed ligand-binding site due to movement in domains shown in green and orange colors. Ligand is shown in ball and stick representation in blue color

(N) and actuator domain (A) [91]. This case study shows the selectivity gain by bound inhibitor, utilizing the domain flexibility of receptor [94].

2.5 Effect of pH on Binding Affinities

Protonation states of the titratable groups participating in the binding can have significant effect on the binding affinity of the interaction [16]. Waelbroeck [95] presented a model with assumptions that correct ionization state of all active groups is the requisite for binding, and ionization state of non-binding residues does not affect binding to study quantitative effect of pH change on binding affinity of the receptor–ligands interaction. They chose pH dependency of insulin and insulin analogs binding to their cellular receptor to study their model [95].

$$\log(K) = \log(K_{\text{real}}) + \log(R^*/R) + \log(L^*/L) \quad (1)$$

where $\log(K)$ is pH-dependent affinity, $\log(K_{\text{real}})$ is reference affinity, R^*/R is proportions of active and total receptor concentrations, and L^*/L is proportions of active and total ligand concentrations. Their model under given assumptions allowed them to attribute binding affinity change only due to change in proportions of active receptor and hormone with changing pH, and express pH dependence as function of number and ionization constants of active groups. Performing binding affinity measurement experiments at varying pH for different insulin analogs binding to their receptors, and analyzing data with modeled relationship [95]. Waelbroeck [95] detected two active groups responsible for marked pH dependence in the normal pH range and suggested that these groups could either belong to the receptor or common residues among porcine insulin, casiragua insulin, hagfish insulin, and desalanine–desasparagine insulin analogs [95]. This study opens up a field in medically relevant design of insulin.

A pH-dependent catalytic activity through hydrolyzing cleavage of type-1 transmembrane protein amyloid precursor protein (APP) of the β -secretase BACE-1 result amyloidogenesis in Alzheimer's disease has been reported by McCammon and co-workers. Enzymatic activity of the BACE-1 is highly dependent to the pH, with peak activity at pH 4.5, while significantly active in pH ranges 4–5 only [96]. The in silico study using constant pH replica exchange molecular dynamics simulation [97] (CpHMD) showed pH dependence of binding affinity of BACE-1 with its inhibitors [98]. The experimental binding affinity measured at pH 4.5 was taken as reference for in silico binding affinity predictions in pH range 1–12, for different inhibitor-bound BACE-1 complexes. CpHMD simulations enabled authors to study influence of conformational dynamics on the protonation equilibria and thereby pH dependence on binding affinity. The microscopic pK_a values of the aspartyl dyad residues Asp32 and Asp228 in apo- and holo-BACE-1 can be estimated from CpHMD simulation data, and protonation changes were observed in apo- and holo-forms suggesting their thermodynamic linkage. They also studied effect of

protonation equilibria on conformational dynamics for the apo BACE-1 with fixed protonation states for titratable residues using conventional molecular dynamics (cMD) in acidic (pH range 1–3) and basic (pH range 9–11) conditions and observed that in acidic condition, two major conformations open and closed were populated while in basic condition, only widely open flap conformation was significantly populated. In another similar in silico study, again using CpHMD replica exchange simulation Ellis and Shen [96] reported that BACE-1 majorly occupies three conformations (so called Tyr-inhibited, binding-competent, and Gln-inhibited) and conformational population shift with varying pH causes the pH dependence of the inhibitors binding affinity to BACE-1 [96]. They showed that Gln-inhibited and binding-competent conformational states are separated by small (<1 kcal/mol) free energy barrier, and Gln-inhibited state has consistently low population (<25%) for entire pH range; thus, they focused on only remaining two of the conformational states, suggesting that substrate BACE-1 binding follows a conformational selection model [96].

2.6 *Effect of Solvation*

Almost all biological functions occur in cytosol in cell, but some of them are membrane-associated phenomena, water solubility of inhibitors showing significant binding affinity toward its cognate receptor poses another challenge in SBDD [99], since low solubility causes low bioavailability of the inhibitor to target. Similar problem surfaced with the potent non-peptide cyclic urea analogs of HIV-1 protease inhibitor, e.g., DMP-323, the carbonyl oxygen of cyclic urea of DMP-323 mimics a structural water in the binding site by providing similar hydrogen-binding features and therefore gains affinity by displacing the water. The low-molecular-weight compound was expected to have high bioavailability [100], but unexpectedly low bioavailability was observed later on, and poor solubility of DMP-323 in water and lipid milieu was suggested the reason for it [99]. Therefore, to increase water solubility, benzylic-substituted cyclic urea with strong acid or basic groups were designed, but highly basic group analogs were unsuccessful as inhibitory effect of such compounds is lowered by 1000-fold [99]. However, a neutral form binding, weak-basic derivative bis-meta-aminobenzyl, i.e., DMP-450 showed enhanced affinity. DMP-450 has enhanced water solubility and also found to show better oral bioavailability in animal species, rat and human [99].

2.7 *Covalent Inhibitors*

Non-covalent inhibitors bind to the target reversibly in concentration dependent manner. However, ~30% of FDA approved drugs are covalent binders, which make covalent bond with the target [101]. Aspirin induces irreversible acetylation

of a serine residue (Ser516) in the cyclooxygenase site of the human prostaglandin endoperoxide H synthase-2 (hPGHS-2) [102], β -lactam antibacterials forms covalent bond with the active site serine of penicillin-binding proteins which inhibits cell wall synthesis of bacteria and causes its death, and tetrahydrolipstatin a fat absorption inhibitors acts by inhibiting activity of pancreatic lipase [103]; these are among the blockbuster drugs and examples of covalent inhibitors. Although non-covalent docking is more common, recently resurgence of covalent docking has been observed [101]. The covalent docking is more complicated mainly because their action between receptor and ligand has to be taken care of. Selectivity of the inhibitor toward target is important to avoid cross-reactivity. However, selective targeting via ligands equipped with different warheads makes covalent inhibition important [104]. In covalent inhibition, an electrophilic ligand binds to a nucleophilic target receptor via forming a covalent bond. Theory and application aspects of covalent docking have been reviewed elsewhere [101]. A comparative study of recent methods and tools, e.g., CovDock [105], AutoDock4 [106], FITTED [107], MOE [108], ICP-Pro [109], and GOLD [110] for covalent docking has also been recently published [104].

2.8 *Functionally Relevant Structure*

Biologically important molecules are involved in very diverse functions and possess the structural, modular, and interactional diversity to carry their functions in the cell. Numerous enzymes are monomer, while several of them are functional only as homo-/hetero-multimeric forms, e.g., PNP is a homotrimer [29], HIV-1 protease is a homodimer but has slight difference in structural features of the two monomers [90]. A large number of macromolecules catalyze enzymatic reactions, e.g., BACE-1 is responsible for catalyzing hydrolytic cleavage of amyloid precursor protein (APP) [111], some of them modulate their functions, e.g., MDMX/MDM2 complex suppresses activity of p53 and activate p53 pathway in tumor cells [74], some of them regulate, and some of them are not related to enzymatic activities at all, like ion channels and signaling related proteins. When we are designing structure-based drug, we are to face challenges posed by structural, functional, and reactional mechanistic diversity of target molecules as well.

The purine nucleoside phosphorylase (PNP) is a homotrimer and hosts three active sites each near the interface between two monomers, with monomer consisting an α/β -fold formed from a β -sheet of four strands, a β -sheet of six strands forming a distorted barrel, and eight α -helices [34]. The interaction between monomers will influence the binding of ligands.

HIV-1 protease is a homodimer consisting of 198 residues. McCammon and co-workers proposed a terminology to describe the topology as follows: flap (43–58), ear (35–42), cheek (cheek turn = 11–22 and cheek sheet = 59–75), eye (23–30), and nose (6–10) [112]. The active site of HIV-1 protease is covered by β -hairpin flaps of the two monomers and is involved in controlling polypeptides'

access to the active site before binding and closing the active site during the cleavage and then release of the cleaved substrates. The flexibility of the flap plays a crucial role in the catalytic activity of the enzyme [113].

Isocitrate dehydrogenases (ICDs) are another group of interesting enzymes with two isoforms—one NADP⁺-dependent homodimer and another NAD⁺-dependent heterotetrametric isoform consisting of two α -subunits one β -subunit and one γ -subunit. As observed in understanding the mechanism of action during phosphorylation, the structural motions facilitate the flap to cover or open the active site, thus providing two different structures of dimmers; hence, the designing needs to take care of such two state structures of receptor [61].

3 Mapping Interaction at Binding Site

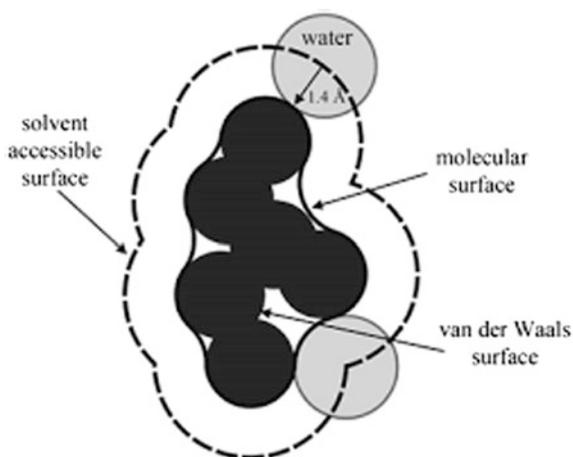
The primary focus of structural biology has been to study the relationship between structure and function of macromolecules. The evolution of protein structure to confer specificity and affinity is still not completely understood. Analysis of related structures has potential to yield local structural regions which are conserved and those which diverge. Such knowledge can potentially be translated into understanding proteins evolution to attain specificity or protein acquiring completely new function by matching curvature along the protein backbone to find structurally active site regions [114].

3.1 Identification of Active Site or Binding Site

The binding sites of most proteins are extremely specific and can determine even very small structural differences among putative binding patterns [114]. Folding of a protein can be considered to be a process which generates specific binding site or cavity from an unstructured polymer, driven and stabilized by thermodynamic forces [115]. Knowledge of protein cavities provide clue about the structure and shape of binding molecule [116]. Ligand-binding sites of protein provide insights to its biological function and reaction mechanism. Identification and application of druggable active sites of target proteins are pivotal in *in silico* drug design [117]. A very diverse active site of a protein is particularly useful for target-based drug discovery as it serves as a prerequisite for protein–ligand docking, which is integral part of structure-based drug design. Accurately predicting the binding modes of inhibitors in the active sites of protein is still observed as a challenge in drug discovery [10].

All the methods which identify the active site of receptor use the concept of accessible surface area as defined by Lee and Richards [118]. The accessible surface (ASA), also known as solvent-accessible surface area (SASA) if water is used as the probe, of a protein is stated as the locus of the center of the solvent molecule as it rolls along the protein, making the maximum permitted van der Waals contacts

Fig. 4 Surface area definition (courtesy: Wikipedia)



without penetrating any other atom. The ASA is closely related to the concept of the solvent-excluded surface also known as the molecular surface or Connolly surface. The cavity identified in protein molecules, effectively the inverse of the solvent-accessible surface, is the binding site as to be used by ligand to satisfy the available physical and chemical interactions. This is pictorially shown in Fig. 4.

Major methods to find the shape of active site using the 3D coordinate of protein or receptors can be classified as approximate and exact method depending on their numerical depth and accuracy in calculation involving the coordinates exclusively. Most of the approximate methods rely on numerical integration where some of them are analytical [119]. Connolly in 1983 [120] introduced the exact analytical methods for computing the accessible surface area. The computational efficiency and robustness has been improved in recent years, but the reduction in overlapping surfaces remains computationally expensive. The difference between approximate and exact computation is applied to existing methods evident from the detail calculation of the derivatives of the surface area with respect to atomic coordinates. All well-known methods used for computing the active site mapping by surface area suffer from the reproducibility problems. A method called Alpha shape [121] uses Delaunay triangulations and computes the surface area and volume of proteins as well as detects and measures cavities in proteins, as described by Edelsbrunner [122], to reduce the overlap. The Alpha shapes method employs a precision geometric method called triangulation to evade numerical problems by systematically resolving all singularities without explicitly perturbing positions of centers of spheres [123]. To provide fast calculation, an extension of the Alpha shapes method that includes the efficient, robust, and exact analytical computation of the derivatives of surface area terms has also been worked out [124].

Based on shape and ASA, many Web-based and stand-alone software are available as listed in Table 3 to find cavity and identify active site of known protein structures.

Table 3 A list of some popular Web servers and stand-alone tools based on shape and ASA formalisms

SN	Programs	Based	Web site links
1	CASTp [125]	Web	http://sts.bioe.uic.edu/castp/index.html?2cpk
2	CCCPP [126]	Desktop	http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html#CCCPP
3	LIGSITE ^{csc} [127]	Web	http://projects.biotec.tu-dresden.de/pocket/
4	KVFinder [128]	Desktop	http://lnbio.cnpem.br/facilities/bioinformatics/software-2/
5	PASS [129]	Web	http://www.ccl.net/cca/software/UNIX/pass/overview.shtml
6	PrinCCes [130]	Desktop	http://scholar.semmelweis.hu/czirjakgabor/s/princces-download/#t1
7	POCASA [131]	Web	http://altair.sci.hokudai.ac.jp/g6/Research/POCASA_e.html
8	RosettaHoles [132]	Desktop	https://www.rosettacommons.org/
9	SURFNET [133]	Desktop	http://www.cgl.ucsf.edu/chimera/current/docs/ContributedSoftware/surfnet/surfnet.html
10	VOIDOO [134]	Desktop	http://xray.bmc.uu.se/usf/voidoo.html

3.2 Characterization of Active Site

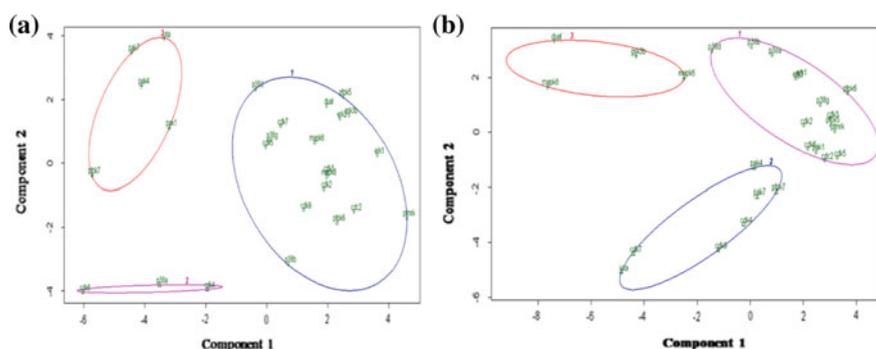
Identification of active sites in large binding pockets in protein or macromolecules does not assure the correct or native poses of ligand binding because many subsites interaction influence the binding of ligands, which has been exploited favorably in case of designing combinatorial ligands of monoamine G-protein coupled receptors (GPCRs) [135]. To design a ligand which effectively come out to be a functional inhibitor requires prior knowledge of interacting subsites and their role to k_{on}/k_{off} kinetics of binding, which until recently [136, 137] hardly have been explored. Our study using kinases, from *P. falciparum* and from human, shows the selectivity of subsite also residing in active site [138]. Using ser/thr kinase sequences of human and plasmodial species those having PDB structure, a phylogenic tree was constructed. Human kinase proteins (22 of them having structural superimpossibility $<2 \text{ \AA}$ RMSD of main chain atoms) shown in Table 4 are listed by sequence as

Table 4 Binding site clustering using sequence of human and plasmodial ser/thr kinase

Plasmodial kinases	Neighboring human kinases
Pfpk5	h_CDK4, h_CDK5, h_CDK3, h_CDK2, h_CDC2
Pfpk6 Pfmrk	h_CDKL1, h_CDKL4, h_CCRK, h_p38a, h_p38b, h_ERK1, h_ERK2, h_CDK10, h_p38d, h_CDK6, h_CDK7, h_p38g, h_CDK9
Pfpk7	h_SmMLCK, h_NEK1, h_LATS1, h_LATS2

Table 5 Selective binding site clustering using structure of human and plasmodial ser/thr kinase, uncommon one shown in bold face font and underlined

Plasmodial	Human		
	Kinase domain	ATP-binding site	Substrate-binding site
Pf $\text{p}k5$ Pf $\text{p}k6$ Pf $\text{m}r\text{k}$	CDK5, CDC2, CDK3, CDK9, ERK2, ERK1, p38- γ , p38- β , GSK3- β , DYRK1A, MAPK8	p38- δ , CDK5, p38- γ , <u>CDK7</u> , <u>MAPK6</u> , CDK3, ERK2, <u>GSK3-β</u> , <u>MAPK8</u> , CDK2, ERK1, <u>CDK9</u> , p38- β , CDC2, <u>DYRK1A</u>	CDK5, CDK3, ERK2, CDK2, ERK1, p38- γ , p38- β , p38- α , p38- δ , CDC2, <u>CDK6</u> , <u>PAK1</u>
Pf $\text{p}k7$	MAPK6, PAK1, PAK4, PAK7, PKC iota	<u>PAK1</u> , PAK4, PAK7, PKC iota	PAK4, PAK7, <u>CDK9</u> , <u>CDK4</u> , PKC iota, <u>CDK7</u>
Non-plasmodial cluster	CDK2, CDK4, CDK6, p38- α , p38- δ	CDK6, Cdk4, p38- α	DYRK1A, MAPK8, MAPK6, GSK3- β

**Fig. 5** Structure-based clustering of human kinases associated with *plasmodium* using **a** ATP-binding site and **b** using substrate-binding site. It clearly depicts different combinations of selectivity (listed in Table 5)

nearest neighbors of specific plasmodial kinases; their 3D structures are used for finding selectivity profile at the active sites. Separately, the ATP-binding and substrate-binding site domains of these kinases are extracted on the basis of Hunk and Hunter classification [139], and their structures are superimposed for clustering on the basis of RMSD matrix and are shown in Table 5 and Fig. 5.

It is interesting to note that three of the *plasmodium* kinases occur in the largest cluster containing most of human kinase, like MapK and CDKs; but PfPK7 occurs in different cluster in both ATP & substrate specific clustering, it signifies the selective functioning of this kinase. Hence, to achieve selectivity in favor of malarial ligand requires subsite exploitation and using appropriate designing strategy for docking compounds in search of both specific and selective ligand. In a recent review [39], such small active site differences are discussed under the context of how the entropy and enthalpy balances are carried out in free energy estimation

in case of HIV proteases binding with ligands that differ by single functional group, by Freire et al. [140]. It may also happen that all available features in the active sites are not satisfied or they may be satisfied by different orientations or conformation of complementary features in ligands. Hence, it is imperative to have prior knowledge of biological function of active site of receptor and detail mapping/association of the subsites with different functional groups in ligand, before starting the docking of large number of ligands to evaluate the binding competency.

Using cliques of favorable interaction points at active site, emerging from probes of different chemical features among a class of protein, specificity pharmacophore has been generated [141, 142].

This novel method provides a complementary map of a class of active sites for designing new chemical entities which specific as well as selective for the receptors. Figure 6 provides an expanded series of such pharmacophores designed from four plasmepsins, acid proteases of *plasmodium*. Using such tools, designing of ligands is possible which can satisfy all the complementary features available in active sites; this can be used to design compounds with better binding capacity. This method can be applied to design the pharmacophores in search of novel inhibitor

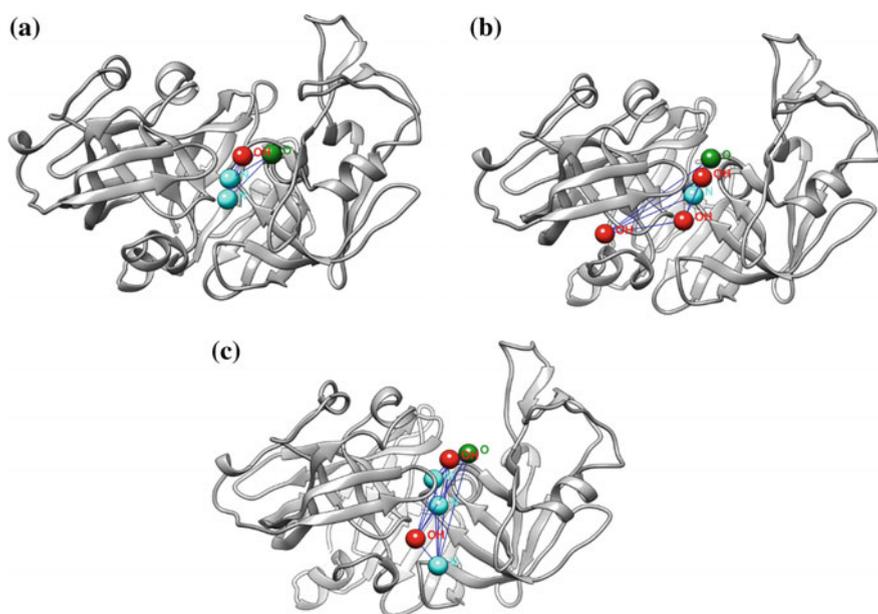


Fig. 6 Utilization of binding site information of class of aspartic protease (human cathepsin, pepsin proteases, and four *plasmodium* plasmepsins) for development of de novo pharmacophore features using in-house program CliquePharm. **a** Four-point, **b** five-point, **c** six-point pharmacophore features, all are shown in cavity of *plasmodium* plasmepsin II (PDB: 1SME), respectively. Nodes are shown as spheres with amide probe in cyan, hydroxyl probe in red, carbonyl probe in green, respectively, and edges are connected

for the multitarget structure-based designing like bacterial multidrug efflux pump and AcrAB-TolC pump [143].

3.3 Why Different Poses?

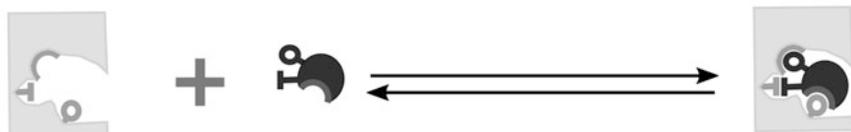
While docking of different chemical ligand at the known active site, one can generate different orientation for the same ligand, which is defined as “pose” due to the fact that many features available in the active sites may or may not be satisfied by the complementary features available in docked ligand. Such variations in interaction between protein and ligand may also occur due to the flexibility of active site residues [14].

Lock-and-key: The lock-and-key model of enzyme substrate interaction proposed by Emil Fisher in 1894. It assumes enzyme-binding site as a cavity with specific set of shape and physicochemical interaction features analogous to the key-hole of a lock, while ligands are potential molecules which possess shape and interaction feature of key, i.e., complementarity [39]. Generally, receptor–ligand interactions are considered to imitate this model during binding. This model was the early motivation for development of docking and scoring studies. However, many interactions associated with the flexibility of ligand upon binding to receptors and vice versa; hence, other models are proposed [36].

Induced fit: The idea of induced fit model (Fig. 7) of binding occurred as many cases the binding site of the protein undergoes subtle arrangements of key residues side chain orientations or conformational changes sensing the presence of ligand in the vicinity under the influence of its interaction fields [144]. For example, drug-target aldose reductase undergoes large conformation change during binding of ligand [79]. Several other cases of this model of ligand–receptor binding are discussed in the section Protein Flexibility.

Conformational selection: This model proposes that the receptor maintains an ensemble of conformations in equilibrium, rather than being into some particular conformational state before binding (as in lock-and-key) or changing conformation sensing the ligand (as in induced fit), whereas ligand binds to the conformation presenting best complementarity at the binding site [39]. For example, BACE-1 binding to and showing significant activity only at narrow pH range 4–5 is actually in equilibrium of at least three Tyr-inhibited, binding-competent and Gln-inhibited significant conformations [96]. However, only binding-competent conformation being conformationally compatible for binding has the highest population at the specified activity pH range 4–5, but the population of these conformation at pH < 4 or pH > 5 is decreased and hence the activity [96]. Another model known as conformational isomerism is found in the literature [14] and has been a special case of the conformational selection, where one or more conformational isomers of the receptor exist in equilibrium and ligand binds to only conformationally compatible isomeric form of the receptor, and binding shifts the conformational equilibrium in

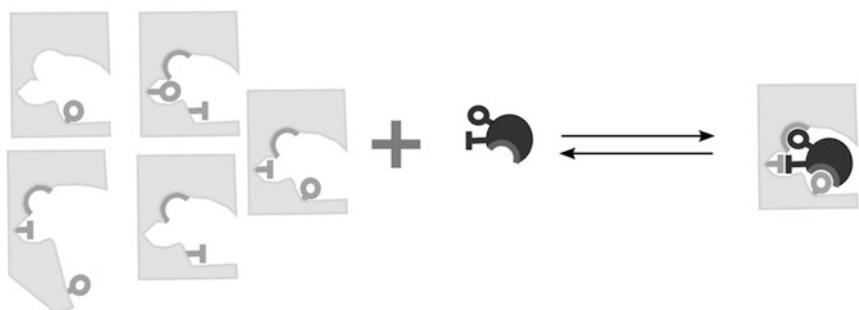
Lock and key



Induced fit



Conformational selection



 Hydrogen bond donor
 Hydrogen bond acceptor
 Hydrophobic

 Ligand
 Receptor

Fig. 7 Schematic representation of enzyme substrate-binding models. Ligand is shown in black color and receptor in gray color. Different binding site features/ligand features are described at the bottom

the direction to establish the equilibrium among conformational isomers. Earlier reported that binding to Fab antibody and catalysis of substrate is restricted to one of the conformation and not to others [145]. In recent paper [146], enzyme catalysis has been prescribed due to conformational dynamics of enzyme active site.

Prediction of poses of ligand with receptor from docking study may differ due to several reasons. For example, model of enzyme action (lock-and-key/induced fit/conformational selection) assumed for the study may not be appropriate to capture the underlying binding mechanism, e.g., assuming lock-and-key for an actual

induced fit or conformational selection case [14]. Existence of possibly alternate interaction features in binding site could provide complementarity for even structurally very similar ligands but provide different poses; several such cases have been reviewed by Teague et al. [147]. Another case could be enthalpy–entropy compensation due to receptor–ligand flexibility for different poses of ligand [147]. Although docking and scoring lack capability to account entropy, considering receptor–ligand flexibility in docking can be a poor proxy for entropy to certain extents.

3.4 Flexibility of Ligand Provides Complementarity

Generally, small molecules can adopt a number of conformations within few kcal/mol energy gap from the global minimum conformation. Thus, a number of conformations of ligands are generated and docked into the receptor to seek optimal complementarity between receptor-binding site and the ligand conformation to yield most probable pose. Therefore, several conformation generation schemes which can be broadly put in two groups, (a) systematic search and (b) random search, have been suggested and are routinely employed in docking studies [148]. Systematic search tries to generate all the conformation corresponding to the rotational states for the rotatable bonds of the molecule, but exponential increase of the number of conformations of the molecule with number of rotatable bonds turns out to be limiting for most of the practical uses. Random search tries to generate different ligand conformations using randomized schemes like genetic algorithm [14, 149].

Small-molecule ligands often interact with binding site presenting complementary features [150]. However, small size of such ligands at times has limited possibilities to interact with neutral binding pockets, because neutral binding site has weak electrostatics interactions and hydrogen bonding capabilities [151]. Neutral and wide open hydrophobic pockets can not present interactions strong enough to portray desired high affinity for small-molecule ligands. On the other hand, peptide ligands due to their flexibility can adopt a wide range of conformations to gain higher affinity in such cases by making more hydrogen bond interactions and through many weak hydrophobic interaction from several hot spots in the pocket [151, 152].

3.5 Is Estimate of Binding Affinity Sufficient?

In case of receptor binding processes, the stability of the binding is accounted by difference of Gibbs free energy between bound and unbound states. The equilibrium dissociation constant K_d which is ratio of unbinding process k_{off} and binding process k_{on} is associated with thermodynamic properties of the reactants/product, whereas the activation energy for the process influenced by kinetic properties [153].

Thus, *in silico* calculated affinity of receptor–ligand binding contains information about thermodynamic parameters and does not include kinetic parameters. All the methods aiming to measure/predict binding affinity would miss the kinetic aspect of the reaction. The kinetic aspect of the process is related to diffusion of the solute molecules under influence of the entropy of the system. Collision of the receptor molecule with the ligand is the requisite for the process to happen. Bigger solute molecules collide with small water molecules and undergo random Brownian motion, and their encounter allows reaction to happen [153].

The dissociation constant K_d represents the ligand concentration in which half of the protein binding pockets are occupied and relate to Gibbs free energy [154] by $\Delta G = RT \ln(K_d)$. Gibbs free energy is a state function and does not depend on the thermodynamic path followed during reaction; it only depends on the initial/final chemical potential of the reactants/products [154]. Association and dissociation rates k_{on} and k_{off} depend on transition states encountered on the pathway during the chemical reaction. Specifically, they depend on highest free energy barrier for the transition state that separates bound and unbound states [154].

Even if the reasonable accuracy in predicting affinity is achieved, it is not sufficient to characterize the protein–ligand-binding process completely [154]. Kinetic aspect of the process can be modeled by mimicking the protein–ligand diffusional encounter in the solvent under thermal fluctuation, which will be discussed later [155].

4 Estimation of Interactions

Scoring functions aim to predict the interaction energy between the receptor and the ligand in a given conformational pose, by summation of weighted interaction features. Scoring functions required to rank chemicals implemented in various docking tools use different assumptions to evaluate modeled complexes [8]. Simplification is achieved at the cost of neglecting full domain flexibility, entropic effects, and solvation effect [8].

4.1 *Different Types of Scoring Functions*

In the literature, wide choice of scoring function is available which can be classified as force-field-based scoring functions, empirical scoring functions, knowledge-based scoring functions, and descriptor-based scoring functions [156]. Force-field-driven scoring functions are based on the molecular mechanics and utilize atomic properties like atomic charge and vdW forces which are already parameterized such as AMBER [157] or CHARMM [158]. Dock6 [159], AutoDock [160], G-score [161], and GOLD [110] are a few popular ones in this class. In scoring functions, only intermolecular interactions are modeled, vdW interactions are expressed using Lennard-Jones

potential function, and electrostatics interactions are calculated using Coulombic formulation. Empirical scoring function [162] on other hand is based on the available physicochemical properties which corresponding to hydrogen bonding interactions, hydrophobic interactions, entropic changes, and interactions with metal ions [162]. Binding free energy is estimated using the sum of various uncorrelated (sometime parameterized) terms derived from the regression analysis using experimentally determined binding energies from the already known crystallized complex structures [163]. ChemScore [163], LUDI [164], Glide score [165], X-score [166], etc., are major tools implemented with such empirical scoring function. Knowledge-based scoring functions [167] are derived from the crystallized protein–ligand complexes using statistical regression principles. The binding free energy of the complex is assumed to be the sum of free energies (potentials of mean force) of interatomic contacts calculated from the frequencies of these interatomic distances in a database of experimental structures from statistical methods [168]. As compared to empirical scoring function, knowledge-based potential function does not require known binding affinity and so are free to explore large and diverse structural complex information to derive the more accurate and less biased scoring function parameters. These functions are expected readily transferred to systems that have not been used in the development of the scoring function. Examples of knowledge-based scoring functions include PMF [169] and DrugScore [170].

4.2 *Nonlinear Relation Between IC₅₀ and Score Values*

A standard scoring function is given in kJ/mol by Eq. 2.

$$\Delta G = 5.4 \Delta G_0 - 4.7 \Delta G_{\text{HB}} - 8.3 \Delta G_{\text{ionic}} - 0.17 \Delta G_{\text{lipo}} + 1.4 \Delta G_{\text{flex/rot}} \quad (2)$$

Assumed to be linear, where coefficients present the weightage of each contribution as mentioned by suffix, in a case study out of 45 known ligand receptors from PDB, the standard deviation having +7.9 kJ/mol or 1.4 log unit error in binding constant. But this is not reflecting reality, which has been observed while comparison of actual and predicted values of binding across the range of activity. Correlation between the binding energies predicted by the docking programs like AutoDock, GOLD and FlexX [171–173] with the experimentally determined binding free energies is analyzed among a set of known ligands in the literature [110, 174]. Prediction of affinity using scoring function has been used for ranking compounds, while high-throughput screening but compared with known experimental data it has been observed that high-affinity compounds (\sim nM) are predicted with lower errors than weak binders (μ M to mM). Generally, the weak binders are overpredicted, whereas tight binders (μ M) are underpredicted [171, 175]. It may require implementing functions to address negative co-operativity so that present scoring functions are trained to penalize weak binding. Tight binders required to be associated with positive co-operativity. However, a measurement of applicability is

done using reproduction of geometry from complexed crystal structure, comparing close relation with binding affinity (experimental) and scoring and ranking and two other important parameters known as enrichment factors (EF) and receiving operators characteristic (ROC) [176–178].

4.3 Does Scoring Function Reflect Binding Activity?

Scoring functions can only predict the binding affinity of a receptor with its ligands in isolation [156], but the cellular environment is significantly different, where it may be interacting with other molecules which may alter its affinity toward its ligand, e.g., activation of tumor suppressor protein p53 activation is regulated by MDM2/MDMX [74]. Inhibition measure of a ligand for its receptor is the end result of several pharmacokinetic factors as well other than affinity, e.g., bioavailability [73, 99]. Therefore, docking score of a ligand for its receptor may not be the actual measure of its inhibitory potential always. The similar kind of evidence emerged, when it was noticed that urea analog DMP-323 had shown good affinity and predicted inhibitory potential for HIV-1 protease [100], but it could not succeed because of its very low bioavailability due to its poor solubility [99]. In the follow-up study, a new analog DMP-450 with higher water solubility was designed and found to show better inhibition of HIV-1 protease [99]. As detailed in Sect. 2.3, in the similar way to save from proteolytic cleavage, α -helical clipped peptide was designed from human serum protein HSP's variants, as inhibitor of the MDM2 and MDMX complex [73]. The proteolytic cleavage was hampering its bioavailability; thus, clipped α -helical peptide achieved improved pharmacokinetics, thus ensured better efficacy in human and rat models [73].

5 Limitations of Methods

5.1 Appropriate Structure of Receptor to Select

While selecting a receptor structure for initiating docking study, parameters listed in Table 1 can be used to prioritize structures if more than one structure is available, and to choose appropriate structure. In present case, we have summarized some of the structure validation results for two different structures of HIV-1 protease (PDB ids: 1FQX and 4ZIP) in Fig. 8 and crystal structure details shown in Table 6. Analysis of structures is available from RSCB PDB [179] (https://files.rcsb.org/pub/pdb/validation_reports/fq/1fqx/1fqx_full_validation.pdf and https://files.rcsb.org/pub/pdb/validation_reports/zi/4zip/4zip_full_validation.pdf).

In general, structure for which different parameter values are in blue zone in horizontal bars for it is preferable. These horizontal bars represent statistical likelihood of reported structure to be in acceptable/unacceptable range. The range of

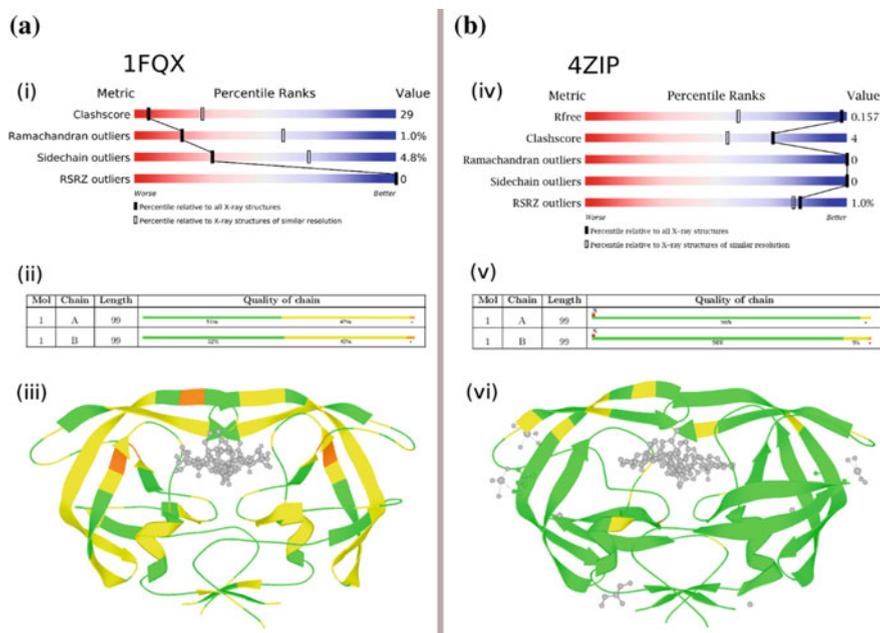


Fig. 8 Two crystal structures of HIV-1 protease shown **a** 1FQX.pdb and **b** 4ZIP. (i) and (iv) show structure quality summary obtained from RCSB Protein Data Bank (PDB). (ii) and (v) show conformance to geometric quality criterion of model residues: 0, 1, 2, and ≥ 3 geometric quality criterion outliers are shown in green, yellow, orange, and red colors, respectively. (iii) and (vi) show mapping of model validation results with electron density over 3D structure for PDBs 1FQX.pdb and 4ZIP.pdb, respectively

Table 6 Crystal structure parameters for HIV-1 protease structures with RCSB PDB (www.pdb.com) codes 1FQX and 4ZIP

Parameter	1FQX	4ZIP
Resolution range low ^a	26.00	50
Resolution high ^b	3.1	1.11
Completeness	Not available	91.7%
R_{work}	0.180	0.130
R_{free}	Not available	0.154
RMSD (bond lengths)	0.080	0.015

^aA minimum spacing (d) of crystal lattice planes that still provide measurable diffraction of X-ray

^bAdditionally, $\langle I/\sigma(I) \rangle$ greater than 2 in high-resolution shell

parameter value is determined from all the structures already deposited in PDB of similar resolution range. As we see from the report that clash score, Ramachandran outliers and side chain outliers' values are higher than acceptable and are in red zones (statistically unfavorable) of their respective bars [180] for 1FQX. While in the case of 4ZIP, all the parameter values are in the blue zone (statistically

favorable). Again, when looking at the geometric quality criterion for two structures, 1FQX only (chain A: 51% and chain B: 52% residues) does not have any outlier, while rest (chain A: 47% and chain B: 40%) have at least one outlier. The geometric quality for 4ZIP seems better as in this case 96 and 90% residues (chains A and B, respectively) do not have any geometric outlier. Further considering fit quality of the model to electron density, 1FQX has certain residues which has at least two outliers and a significant percentage of residues with at least one outlier, while in case of 4ZIP, there are no residues which have two outliers and only a small fraction of residues with only single outlier. Considering all above points among 1FQX and 4ZIP, 4ZIP should be preferable over 1FQX as receptor structure for any docking study.

In Fig. 9, the docking using Dock6 of ligand GRL-0648A to two different receptor structures of HIV-1 protease (4ZIP: high resolution and 1FQX: low resolution) is performed to assess the effect of receptor structure quality on outcome. Results show that when ligand was docked to native receptor structure (4ZIP), it reproduces the crystallized pose (RMSD: 0.40 Å, see Fig. 9a), with dock score of approximately -125. When we docked ligand to poor receptor structure (1FQX), it docked in different poses where core group adopts similar pose but the 5-atom ring (1 nitrogen, one oxygen) containing methyl adopts different poses and leans over Gly48 on chain B, score is significantly low (-14) and RMSD: 2.71 Å (Fig. 9b). This observation suggests that high-quality receptor structures are more likely to present better interaction complementarity, saving from predicting high-affinity binders mistakenly as poor-affinity ligand.

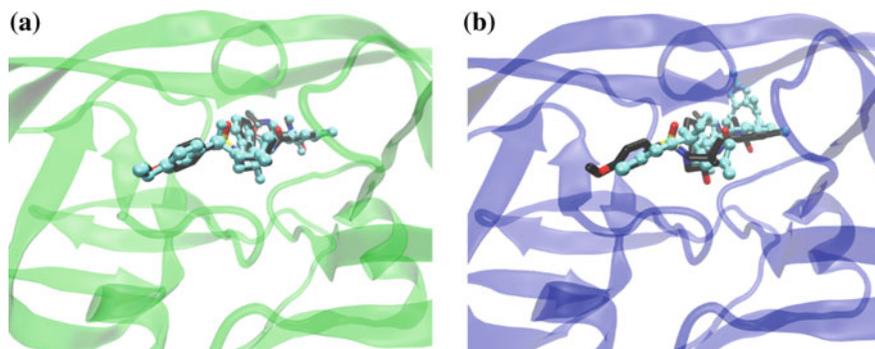


Fig. 9 HIV-1 protease-binding site structures shown **a** HIV-1 protease structure (PDB: 4ZIP) in complex with GRL-0648A (isophthalamide-derived P2-ligand), receptor-binding site is shown in green ribbon and crystallized pose of GRL-0648A in black stick. GRL-0648A is docked to the receptor using Dock6 and docked pose is shown with ball and stick representation and carbons colored in cyan, RMSD of docked pose with reference to crystallized pose is 0.40 Å over 49 non-hydrogen atoms. **b** HIV-1 protease structure (PDB: 1FQX) with GRL-0648A crystallized pose (taken from 4ZIP after superimposing receptor structures) shown in black stick, docked pose of GRL-0648A shown in ball and stick representation with carbons in cyan color, docked pose RMSD 2.71 Å over 49 non-hydrogen atoms

5.2 Analysis of Docking Tools

As discussed above, it is fruitful to analyze the ligands binding efficiency using many methods like AutoDock, GOLD, Glide, LibDock, and HADDOCK; all these tools are different in the method of docking as well as scoring.

There are several open-source commercial but free for academic use, and complete commercial docking programs available from different software vendors. In particular, fifty-one stand-alone and nineteen Web servers for docking employing diverse set of novel features are listed at <http://www.click2drug.org/index.html#Docking> (accessed on Dec 2017). To select suitable program(s) for docking studies for receptor(s) of interest requires insight and expertise [117] in the method. However, we shall discuss only a few selectively chosen methods based on popularity and diversity of strategies implemented in them as shown in Table 7.

Here we are discussing the in-house case study (unpublished work) of four docking programs used to dock already experimentally known inhibitors of *P. falciparum* protein kinase 5 (PfPK5) with IC₅₀ values ranging from 130 to 15000 nM. PfPK5 is a ser/thr kinase and homolog of human CDK2 [185]. Chosen inhibitors are olomoucine (OLM), indirubin-5-sulfonate (INR), staurosporine (STA), and purvalanol B (PVB), respectively. Crystal structures of two of the inhibitors (INR and PVB) in complex with PfPK5 are available [185]. We have chosen LibDock v2.3, Gold v5.2, Dock v6.7, and Glide v7.0 for the comparison study. Different docking programs use different scoring schemes, e.g., Glide score and Dock score assign high negative score to high-affinity ligands, while LibDock and Gold assign high positive score to high-affinity ligands. Pose reproduction and also scoring/ranking of docked poses of these inhibitors is a good case to assess comparative performance of each of the selected docking program and also with experimental values.

The best-scoring poses predicted by each of the programs were compared with the crystallized poses for selected available complex of PfPK5 with PVB as in PDB (1VOP). Predicted poses for PVB obtained from LibDock, Gold, Dock, and Glide showed 0.60, 1.01, 0.88, and 1.87 Å RMSDs with crystallized pose, respectively. In present case, all the selected programs were able to reproduce observed binding mode within RMSD of 2 Å.

Docking and scoring results obtained from the chosen programs show that none of these could predict the correct ranking against the experimentally known activity of chosen inhibitors (see Table 8). The best binder (PVB) among four inhibitors is predicted to be best binder as rank 1, by Gold and Dock6, while LibDock and Glide have ranked 2. LibDock is unable to discriminate between the OLM and INR and predicts them as rank 3 and rank 4, while experimentally found ranks would be 4 and 3, respectively. Again, LibDock does not discriminate between STA and PVB and predicted ranks are opposite to the experimental ranks. Gold predicts correct ranks for best and worst inhibitors, while is unable to discriminate between mid-ranged inhibitors INR and STA. Dock6 predicts correct ranks for better binders STA and PVB, while does not discriminate between weak binders OLM and INR. Predicted ranks from Glide did not match with experimental rank for any of the four

Table 7 Summary of docking approach used, techniques for ligand and/or receptor flexibility, and major features available in some chosen popular docking programs

Programs	Ligand flexibility	Receptor flexibility	Major features in brief
AutoDock [181]	Genetic algorithm	modeling flexible residues	Force field-based scoring function, uses averaged interaction energy grid to account for receptor conformations and simulated annealing for ligand conformations
DOCK [159]	Incremental build	Yes (through AMBER score)	Force field- and contact score-based scoring functions; docks either small molecules or fragments, include solvent effects
Glide [165]	Exhaustive search	No	Empirical score. Although, receptor flexibility can be used in Induced fit, Docking (IFD workflow) with Glide and side chain rotations through PRIME
GOLD [110]	Genetic algorithm	Side chain flexibility and ensemble docking	Empirical score, highly configurable allowing to utilize chemical intuition and domain expertise to improve pose prediction and virtual screening
HADDOCK [182]	Yes	Semi-flexible torsion angle refinement	Uses biochemical and/or biophysical interaction data such as chemical shift perturbation data resulting from NMR titration experiments, mutagenesis data, or bioinformatic predictions
LibDock [183]	Rigid docking can use programs in suit to generate conformation	No	Docks a pre-generated set of conformations for the ligand followed by a final flexible gradient-based optimization of the ligand in the protein binding site
LigandFit [184]	Monte Carlo	No	Empirical score, ligand conformation docked into an active site based on shape, followed by further CHARMM minimization

Table 8 Summary of docking scoring/ranking results of chosen four inhibitors with known IC₅₀ values to PfPK5

Inhibitor	IC ₅₀ (in nM)	RT ln(IC ₅₀) (kcal/mol)	Docking score			
			LibDock ^a	Gold ^a	Dock6 ^b	Glide ^b
OLM	15,000	-6.622	107.08(3)	55.56(4)	-56.46(3)	-5.75(3)
INR	5,500	-7.220	106.80(4)	64.56(2)	-55.04(4)	-8.65(1)
STA	1,000	-8.236	132.82(1)	60.98(3)	-64.55(2)	-4.99(4)
PVB	130	-9.453	130.25(2)	78.40(1)	-71.41(1)	-7.98(2)

Docking score from programs is given in cells of table, while rank is given in pair of parentheses. Four docking programs, LibDock v2.3, Dock v6.7, Glide v7.0, and Gold v5.2, were used to dock inhibitors in the PVB bound structure of PfPK5, after removing PVB. Inhibitors are tabulated from top to bottom in increasing affinity order

^aHigher positive score represents higher affinity

^bHigher negative score represents higher affinity

inhibitors. A limited study like this brings out the uncertainty in pose and rank prediction by popular tools.

5.3 Selection of Appropriate Database

Chemical databases are selected from the ensemble of the small organic and synthetic molecules, used for ligand docking, constituents of such chemical libraries influence the final outcome in the drug designing process. In general, chemical library databases are created to aid the drug discovery process by providing innovation in new lead structures selection. After the establishment of the in silico drug designing protocol, chemical databases are screened to identify the probable inhibitors which can be tested by experimental methods. Success rate in finding true inhibitor by in silico means depends upon both screening protocol and chemical databases used. So, before the selection of the chemical libraries, basic biological target specific chemical features should be marked. For the virtual screening purposes, the compound database may be selected in such a manner so that maximum structurally diverse chemicals can be utilized against the studied biological target(s). Chemo-informatics tools are mainly used not only for diversity analysis [186, 187] but also for converting them into focused chemical libraries [188].

Various chemical compounds databases are available which include databases of general organic compounds intended for screening, drugs, commercial databases, and databases with known biological activity, crystal structure information, and various physicochemical properties information [189, 190]. Table 9 shows some of the commonly used chemical databases which are categorized based on the different features like associated bioactivity information, known drug information, and having target specific information. Most of these databases provide chemical information using 1D representatives such as SMILE and InChI Key, or 2D structural coordinate information stored in SD file format. These databases are also provided online interface to access the whole chunk of chemical compounds for similarity-based screening. These functionalities intended to search close analogues of known bioactive compounds and thereby advances the lead optimization process.

Though different chemical databases are available for virtual high-throughput screening (vHTS), it is recommended to convert any chemical library to “target or focus” chemical library to avoid the false hits selection as novel inhibitor [191]. In the literature, several characteristic properties of small molecules have been discussed that are followed by the “lead-or drug-like” molecules and are considered to be important for a drug to be successful [192]. Currently, list of open-source Chemoinformatics tools is available which can be utilized for drug-like properties calculation and chemical databases filtration [193]. Well-known physicochemical properties which are used as empirical rules are Lipinski’s “Rule of Five” [194], “Rule of Three” [195], and Pfizer’s “Rule of 3/75” [196] (Table 10). Apart from filtering for lead-like properties, it is also important to exclude known toxicophores or metabolically liable moieties which can interfere with the assay and detection protocol.

Table 9 Some commonly used chemical databases

Databases	Web link
<i>Bioactivity data</i>	
Binding activity database	https://www.bindingdb.org/
ChEMBL	https://www.ebi.ac.uk/chembl/db/
NCI	https://cactus.nci.nih.gov/download/nci/
PDB bind database	http://sw16.im.med.umich.edu/databases/pdbbind/index.jsp
PubChem	https://pubchem.ncbi.nlm.nih.gov/
<i>Patents</i>	
IBM	www-935.ibm.com/services/us/gbs/bao/siip/
SureChEMBL	www.surechembl.org
<i>Drugs</i>	
DrugBank	www.drugbank.ca
FDA	http://fdasis.nlm.nih.gov/srs/srs.jsp
<i>Available for vHTS</i>	
ZINC	http://zinc.docking.org
ChemSpider	http://www.chemspider.com
eMolecules	www.emolecules.com
MDL drug data report (MDDR)	http://accelrys.com/products/collaborative-science/databases/bioactivity-databases/mddr.html
BioPrint	http://www.cerep.fr/cerep/users/pages/ProductsServices/bioprntservices.asp
<i>Target specific</i>	
Pfaldb	http://pfaldb.jnu.ac.in/Malaria/homeHit.action
Mycobacterium DB	http://tbnetindia.in/
Therapeutic target database	http://bidd.nus.edu.sg/group/cjttd/TTD_HOME.asp
KLIFS	http://klifs.vu-compmedchem.nl/
Kinase profiling inhibitor database	http://www.kinase-screen.mrc.ac.uk/kinase-inhibitors
<i>Structural databases</i>	
Cambridge crystallographic data center	https://www.ccdc.cam.ac.uk/
Crystallography open database	http://www.crystallography.net/cod/

There is a well-recognized need of creating standard datasets for which experimental bioactivity of the ligands is already known for receptors coming from various functional classes [197] in the research community. Availability of standard dataset for benchmarking docking would potentially aid to spot limitations and non-optimal parameter sets used for docking and scoring with the concerned docking program and thereby allowing tracing and possibly fixing of issues in earlier phases of the study. Development of benchmarking datasets for docking and scoring has been reviewed recently [197, 198]. Primary attempts toward docking was made by Bissantz et al., a dataset contained estrogen alpha receptor (ER α) and

thymidine kinase (TK) with one PDB structure, ten active compounds, and 990 randomly selected decoys from pre-curated Advanced Chemical Directory (ACD) which was considered for each of receptors to evaluate DOCK, FlexX, and GOLD programs and seven scoring functions (Dock, FlexX, GOLD, PMF, ChemScore, Fresno, and Score) [197].

5.4 Consensus Evaluation of Docking

Docking studies performed using different programs which do not necessarily agree with each other as discussed earlier, mostly because each program carries different subtasks of docking with potentially different approach [199]. Thus, when results disagree among themselves, then selection of the final compounds to test becomes indecisive. Matthew and co-workers [199] suggested selection of results based on consensus followed by rationalization through physicochemical intuition. As discussed later, such strategies should be projected as standard to increase confidence in docking results and decrease failure rate of docking studies.

Benchmarking of docking studies is very important for unbiased evaluation of various docking methodologies and their implementations in docking programs. To address this issue, Huang et al. [176] conducted a study along with creating a directory of useful decoys (DUD) [176]. They choose total 40 different targets with eight nuclear hormone receptors, nine kinases, three serine proteases, four metalloenzymes, two folate enzymes, and ten other enzymes. The crystal structures of all targets except one kinase (PDGFRb) were available in PDB. They used 2950 ligands, creating 36 physically similar but topologically different decoys for each ligand. Docking was done using DOCK 3.5.54, with flexible ligand and a force-field-based scoring function accounting van der Waals and electrostatics interaction energies corrected for ligand desolvation. Authors reported that for most of the targets, with MDDR (Elsevier MDL, San Leandro CA) databases, enrichment were almost half log better than DUD, which supported their conclusion that generally databases have bias.

Another protocol is known as checking with cross-docking which aims to summarize the overall success of docking study [200], it captures ligands specificity for its cognate receptor at diagonal of the matrix, and off-diagonal entries represent enrichments against off-diagonal targets. The off-diagonal enrichments could also be indicative of promiscuity of the ligand, or the similarity of the off-diagonal targets [201]. The cross-docking performed in the process highlighted striking results that ligands having very good enrichment for their cognate receptor had good enrichments against a few other receptor sets, while ligands with poor enrichment for their receptor had poor enrichment against others [202–204].

Overall, it has been found that the interaction-based classification and estimation of accuracy of poses during docking are in better agreement with the experimental results [205].

5.5 Selection of Suitable Scoring Function

Whether to select just a scoring function or a consensus scoring function? A suitable scoring function has important role to play to extract correct poses while docking. Poses should be evaluated by the docking score or the ranks are better for evaluation of docking; these are critical aspects influencing the final outcome of the docking results. None of the available scoring functions appears to be fit in all cases [206]. James B. Matthew and co-workers performed a study to evaluate performance of four individual scoring functions DOCK, GOLD, PMF, and FlexX and several forms of consensus scores (CScore) derived from them, over a dataset of twelve HIV protease and nine thermolysin complexes with known crystal structure and experimental binding affinity [199]. Since DOCK and GOLD scoring functions were not available in FlexX, they implemented these scoring functions according to their open descriptions in the literature and will be referred by D-SCORE and G-Score. They found that none of the considered scoring functions was consistently good for all active sites [206], but the CScore (consensus score) was better than all individual scoring function [199]. Secondly, they studied these scoring functions for scoring candidate ligand configurations over a set of five known receptor ligand complexes (2-MQPA or NAPAP into thrombin (1ETR and 1DWD), 1-3-phenyllactic acid into carboxypeptidase A (2CTC), 1-deoxynojirimycin into glucoamylase (1DOG), and DANA into neuraminidase (1NSD) each of the ligand was docked to cognate receptor, and top thirty configurations with most favorable FlexX scores were chosen for further study, each of these configurations were scored using D-SCORE, G-SCORE, PMF, rank-score, deprecated rank-sum (rank-sum after leaving out worst rank), worst-best and CScore methods. They found that average scores from several methods are better than individual score [199]. Apart from this, their study highlighted that there could be alternate poses for NAPAP binding in thrombin and DANA in neuraminidase as predicted by FlexX along with crystal structure poses reproduced in Fig. 10a, b respectively.

Table 10 Typical physicochemical properties which are used to filter the chemical databases

Properties	Lead-likeness
Molecular weight (MW)	200–500
Lipophilicity (cLogP)	-4/4.2
H-bond donor	≤ 5
H-bond acceptor	≤ 10
Polar surface area (PSA)	≤ 170 Å ²
Number of rotatable bonds	≤ 10
CACO-2 membrane permeability	≥ 100
Solubility in water (log S)	-5/0.5
Others	Absence of both toxic and reactive fragments

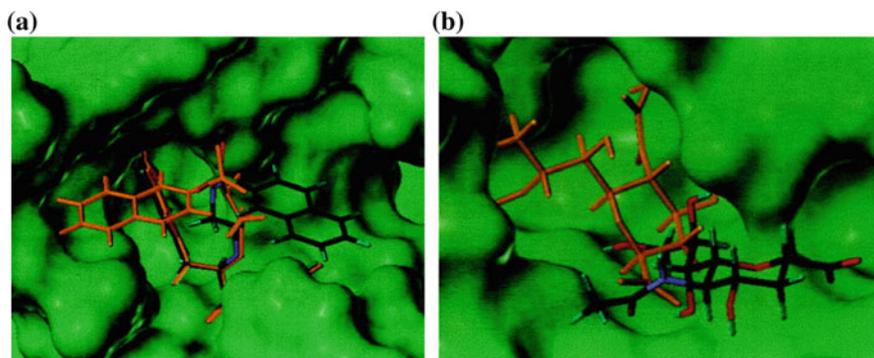


Fig. 10 Alternative docking mode for identified by FlexX and CScore. The alternative configuration is colored by atom type, whereas the binding mode found in the crystal structure is colored orange. **a** NAPAP in thrombin (1DWD) and **b** DANA in neuraminidase (1NSD) [200]. Reproduced with permission

5.6 Consensus Scoring

Despite availability of variety of scoring functions, none of them is universally good for assessment of all receptor ligand binding using docking. Therefore, several attempts [174, 199, 207] have been made by researchers to investigate several scoring functions and their combinations using different consensus schemes. In particular, Oda et al. used two force field-based (Dock score and GOLD score), two knowledge-based (DrugScore and PMF score), and five empirical (FlexX score, ChemScore, PLP, Screen Score, and X-Score) scoring functions and systematically assessed performances of all 511 ($2^9 - 1$) consensus scores over a test set where structures were available in PDB for all chosen 220 protein–ligand complexes. For the sake of comparison, either all the candidate poses scored by a scoring function were ranked assigning best-scoring pose a rank 1 or the scores were scaled to span range 0–1, with best-scoring pose assigned 0 and worst assigned 1. These schemes were consistently used for all the scoring functions, except for X-Score, since it assigns a higher value to better pose in contrast to rest of others. Therefore, S-Score was multiplied by -1 before scaling or ranking [207]. Oda et al. [207] used six different averaging schemes for consensus score with three different ways of model selection (selecting models with consensus score $\leq x_{\text{threshold}}$, top $y_{\text{threshold}}$ models from sorted list of consensus scores in increasing order, and top $z_{\text{threshold}}\%$ models from sorted list of consensus score in increasing order) combined with two ways (by rank and by scaled score) of mapping score to common scale. Prefixes number-by-, rank-by-, and percent-by- were used to denote way of model selection, and suffix rank and number were used to denote ways of mapping scores. Apart from these six, three more double thresholds (one for model selection from $x_{\text{threshold}}$, $y_{\text{threshold}}$, and $z_{\text{threshold}}$ and other number of minimum votes for electing the model)-based vote-by-consensus scores were also evaluated [207]. Considering the

accuracy and efficiency balance in selecting poses rank-by-number and percent-by-number are more useful, while for accuracy number-by-number and vote-by-number approaches are more pertinent to pose selection [207]. GOLD score and Dock score were poor individually but were useful in consensus scoring [207]. Consensus score involving all nine scores or five CScore functions were useful without any optimization and suitable for practical usage [207]. However, Free energy and empirical scoring has been used together in the recent paper [174].

5.7 Inclusion of Flexibility of Ligand and Receptor

In computer-assisted drug discovery process such as structure-based drug design and ligand-based drug design, ligand flexibility plays key role for pharmacophore features extraction and model generation [208], 3D-QSAR analysis [209], molecular docking-based studies [210], shape similarity [211], and so on. In these cases, the outcome results largely depend upon the ability to achieve those conformers that represent the bound state. Hence, it is important to achieve bioactive conformational space of each compounds under study [212]. The term “bioactive conformation generation” specifies the generation of pool of all possible molecular structures that are found in the bound state of the complex macromolecules. Various studies suggest that during the interaction with the receptor, small molecules generally adopt low-energy conformation [213].

The literature suggests two major classes of methods that are utilized to explore the conformational landscape of the small molecules [214]. These approaches include stochastic sampling, systematic or deterministic sampling. Deterministic approaches attempt to generate full range of minimum energy conformations by adopting systematic exhaustively space search approach. This type of space search methods largely dependent upon the number of rotational bonds a small molecule has. Due to combinatorial explosion in torsion angle combinations, this approach is feasible only for very small molecules [214]. Stochastic sampling tries to explore various energy landscapes by incorporating randomness during the search process. Monte Carlo-type (MC) simulations and genetic algorithms (GAs) are the major techniques of this type of sampling methods [214]. A detailed review of these approaches can be found in the following papers [215].

Using above-mentioned approaches, various conformation generation programs have been developed and utilized in drug discovery process cited in Table 11. These programs generally adopt heuristics to overcome combinatorial explosion in case of systematic search and random perturbations and selection in stochastic search.

Ligand being usually smaller in size with lesser number of rotatable bonds exhaustive sampling of available conformational space is achievable with current computational capabilities; but proteins being large macromolecules, available conformational space is vast due to large number of degrees-of-freedom (DOFs) and its exhaustive sampling is almost infeasible. Therefore, techniques seeking to

Table 11 A brief summary of major programs for small-molecule conformation generation

Program	Type	Algorithm	Cost/license	References
Balloon_GA	Stochastic	Genetic algorithm	Free/ proprietary	[216]
CAESAR	Systematic	Incremental search of torsion angles combined with distance geometry	Commercial	[217]
Confgen	Stochastic	Random walk on energy surface	Commercial	[218]
Confab	Systematic	Torsion driving approach	Open source	[219]
Corina	Systematic	Knowledge-based rules derived from CSD	Commercial	[220]
ETKDG	Stochastic	Distance geometry and knowledge base	Open source	[221]
Frog2	Stochastic	Monte Carlo	Open source	[222]
MS-Dock	Systematic	Brute force, anchor, and grow	Open source	[223]
MOE	Stochastic	Random perturbations of rotatable bonds in increments biased around 30°	Commercial	[108]
OMEGA	Systematic	Knowledge-based, complete enumeration	Commercial	[212]
RDKit	Stochastic	Distance geometry	Open source	[224]

incorporate protein flexibility during binding has been attempted, but they incorporate receptor flexibility only to a limited extent, focusing on sampling only most plausible/relevant portion of the conformational space, e.g., through side chain flexibility, conformational relaxation, and multiple structure docking, as already discussed in protein flexibility section. However, newer techniques, e.g., supervised molecular dynamics (SuMD) can be useful to incorporate receptor flexibility, because they allow receptor to experience thermal fluctuation and supervision of ligand toward binding site from unbound state might allow receptor to adopt induced conformational changes sensing the ligand in vicinity of binding site under influence of its interaction field [225].

6 Binding Ability and Free Energy Calculation

The binding free energy of ligand to receptor is the thermodynamic signature of the interaction affinity. Therefore, accurate prediction of binding free energy has been attempted from long times. The free energy calculation methods can be grouped into relative binding free energy calculation methods and absolute binding free energy methods [226]. Relative binding free energy methods aim to calculate

binding free energy of one ligand (reference ligand) relative to another ligand (target ligand) both binding to same receptor, by summing up the work carried to convert one ligand to another in bound and free states in solution [226]. This method can be significantly efficient when reference ligand is very similar to target ligand, but if they are dissimilar then defining and sampling along the conversion path may pose severe computational demand [226]. Since reference ligand to target state conversion path is artificial, these methods are also called alchemical methods, and excellent review on popular methods of this class already exists [227]. Absolute binding free energy methods estimate standard binding free energy of interaction by computing reversible work done in process of transferring it from binding site into solution [226]. Absolute binding free energy methods have been reviewed by Shirts et al. [228]. Practical aspects of free energy calculation have also been recently reviewed [229, 230]. The accuracy of the binding free energy calculations is influenced by adequacy of sampling (theoretically, accurate results require infinite sampling), force field used for sampling, and correctness of the molecular model used, e.g., usually simulation is performed using fixed protonation states of titratable residues, while protonation states might change in experimental conditions [226].

6.1 Calculation of Enthalpy by MM-PBSA

The end-state free energy methods explained here are most common approaches to calculate binding free energy. Linear response approximation (LRA), linear interaction energy (LIE), and molecular mechanics Poisson–Boltzmann surface area (MM-PBSA), molecular mechanics generalized Born surface area (MM-GBSA) [231] are such methods available in the literature. End-state free energy methods are computationally less demanding, but the speed gain in CPU comes at cost of compromised accuracy of the results [231]. These methods are required to be plugged with estimation of configurational entropy which usually is obtained by rigid-rotor approximation and normal mode analysis or quasi-harmonic analysis to yield binding free energy [232]. However, these methods can be good for evaluating binding enthalpy for ligand–receptor interaction. In MM-PBSA/MM-GBSA approaches (schematically shown in Fig. 11), the binding energy is calculated by taking energy difference of free-form of protein (P), and ligand (L) from protein–ligand complex form (PL) [232].

The free energy of each of the molecular species (say X) can be expressed as sum of their molecular mechanics energy in gas phase $E_{\text{MM}}(X)$, solvation free energy $G_{\text{solv}}(X)$, and entropic part— $TS(X)$. The $E_{\text{MM}}(X)$ contribution can be expressed as sum of bonded, electrostatics, and van der Waals energies, i.e., $E_{\text{MM}}(X) = E_{\text{bond}}(X) + E_{\text{elec}}(X) + E_{\text{vdW}}(X)$ [231]. Similarly, $G_{\text{solv}}(X)$ can be expressed as sum of polar and non-polar contributions $G_{\text{polar}}(X)$ and $G_{\text{non-polar}}(X)$, where $G_{\text{polar}}(X)$ can be accounted using Poisson–Boltzmann or its simplified version

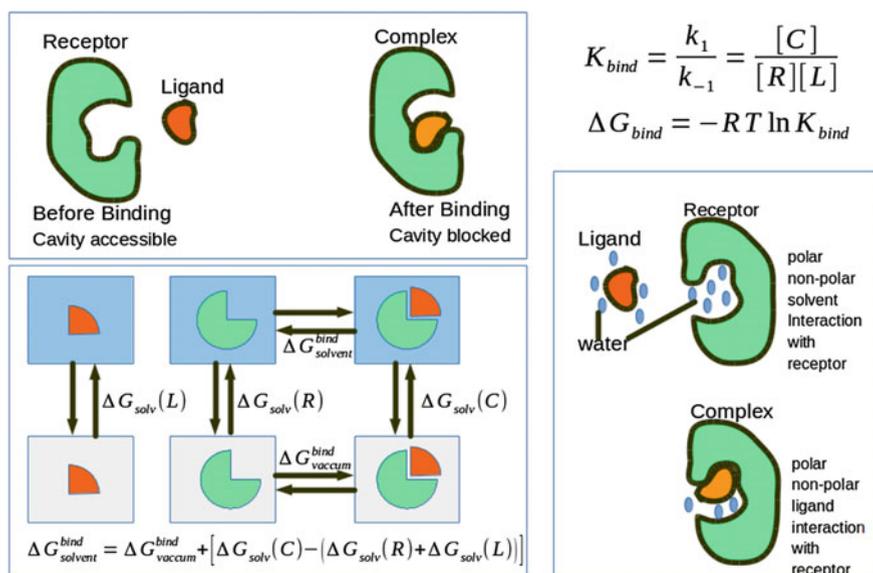


Fig. 11 Schematic representation of the end-state free energy using molecular dynamics Poisson-Boltzmann surface area method for estimating binding energy for receptor ligand binding

generalized Born method as $G_{\text{PB}}(X)$ or $G_{\text{GB}}(X)$, while non-polar is taken to be proportional to accessible surface area change $G_{\text{SASA}}(X)$ [231].

$$\Delta G_{\text{bind}} = G(PL) - G(P) - G(L) \quad (3)$$

The dynamics of the Pfk5 kinase structure complexed with the inhibitor(s) described earlier in docking section is used here as case study using MD simulations. Starting structures of PVB-Pfk5 [185] and INR-Pfk5 [185] complex were taken from crystal structures 1V0P and 1V0O, while OLM-Pfk5 and STA-Pfk5 were taken as consensus pose obtained from docking study using Gold, Glide, and Dock6 as mentioned above. All the systems were prepared using AmberTools14 [233] for MD simulation, and AM1-BCC charges for ligands and GAFF [234] force field parameters with ff14SB [235] parameters for protein. Equilibration was performed using standard protocol [236]. For each case, 12 independent (starting from different starting velocities) MD simulations in NPT ensemble each with length 254 ns were done, initial 4 ns run were discarded to allow for equilibration, bond lengths involving hydrogen were constrained using SHAKE [237] to allow use of 2 fs time step, temperature was controlled using Langevin thermostats with collision frequency 1 ps, and pressure was regulated with Berendsen scheme at target pressure 1 atmosphere using cuda version of program pmemd available in Amber14 [238] MD simulation package. Coordinates were saved every 1 ps. These trajectories were concatenated to yield 3 μs MD simulation for each case containing

3000,000 frames. Every 100th frame was taken for MM-PBSA analysis using MMPBSA.py [239] program in AmberTools14 [233].

The gas phase binding energy ΔE_{MM} was highest for INR followed by STA, PVB, and OLM, but the solvation penalty was also highest for INR and least for STA. In terms of enthalpy of binding, STA was predicted to be best, followed by PVB, INR, and OLM, respectively. The inconsistency of binding enthalpy with IC_{50} indicates possible role of entropy in this case. There may be role of solvation as well which is not rigorously captured in solvation terms considered proportional to buried surface area on binding in MM-PBSA method; see Table 12.

However, it may be criticized that selected docking programs use different scoring, therefore to be able to assess their performance as well as compare with experiment values is not possible. So, another attempt was done by normalizing all the scores, by converting all of them to positive scores (normalized using $(score - \min_{score})/(\max_{score} - \min_{score})$). This yields a consistent normalized score, where weakest and strongest binder ligands get normalized scores ranging 0 and 1, respectively. Same is used for normalizing experimental values, i.e., $RT \ln(IC_{50})$. Results are shown in Fig. 12, Dock6 predicted scores for all ligands are within 1-sigma range, Gold and LibDock each predicted one outlier, and Glide predicted two outlier scores. In present case, Dock6, Gold, and LibDock appear to perform better than Glide. These results may not be sufficient to capture docking/scoring capabilities of chosen programs, as only four ligands are studied and they bind to only one target. A more diverse target set and a large ligand set could better comprehend the features and/or limitation of individual programs; this will be discussed later also.

The binding enthalpy predicted using MM-PBSA method consists of two outliers from 1- σ range (computed as discussed earlier), and it does not agree fully with docking scores obtained from any of the four chosen programs, as expected. However, strong and weak binders predicted using MM-PBSA is same as predicted by LibDock, and second strong binder predicted using these two is similar in affinity. While MM-PBSA results agree with Gold results for two weak binders and not for strong binders. Glide agrees on experimentally found strong and weak binders with MM-PBSA. Score using Dock6 agrees better than MM-PBSA (Fig. 12). As observed in the present case, the scoring by Docking methods as well

Table 12 Enthalpy component of binding free of selected inhibitors of PfPK5, calculated using MM-PBSA method

Inhibitor	IC_{50} (in nM)	RT ln(IC_{50}) (kcal/mol)	Predicted (kcal/mol)		
			ΔE_{MM}	ΔG_{Solv}	Total: ΔH_{PBSA}
OLM	15,000	-6.622	-58.5 ± 7.9	25.4 ± 5.9	-33.1 ± 4.1
INR	5500	-7.220	-102.7 ± 8.7	62.4 ± 5.9	-40.3 ± 4.2
STA	1000	-8.236	-69.0 ± 5.5	19.8 ± 4.0	-49.2 ± 4.7
PVB	130	-9.453	-65.7 ± 8.9	22.6 ± 6.0	-43.1 ± 4.7

These values are computed for 3000 snapshots extracted from 3- μ s-long MD simulations for each inhibitors in complex with PfPK5, internal dielectric constant was taken 2, and ionic strength zero

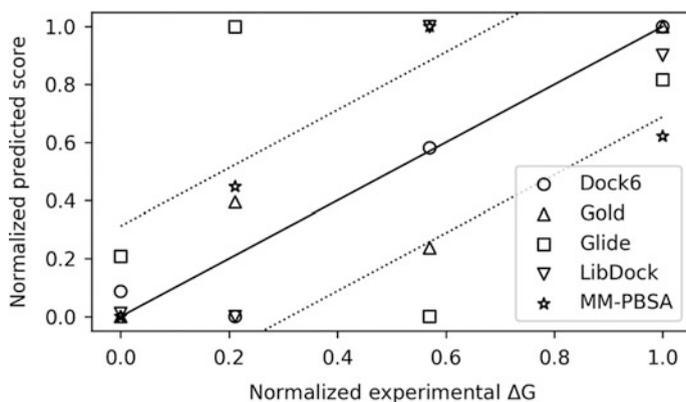


Fig. 12 All the scores have been normalized as discussed in text, to compare the predicted affinities for chosen four inhibitors of PfPK5 obtained using docking with Dock6, Gold, Glide and LibDock and MM-PBSA against experimental binding affinity. Solid line shows perfect correlation of scores with experimental results, and dotted lines above and below show one- σ range of error for predicted affinity

as the end-state Free Energy methods show discrepancies with experimental results, which emphasizes the effect of entropic contribution in case of flexible Kinase binding to ligands.

6.2 Effect of Entropy to Ligand Binding

Gibbs free energy (ΔG) of binding has two components enthalpy (ΔH) and entropy ($-T\Delta S$) as given by Eq. 4:

$$\Delta G = \Delta H - T\Delta S \quad (4)$$

Enthalpy of the protein ligand interaction is assumed to be the major determinant of the binding free energy assuming entropic contributions for smaller ligands binding to the same receptor would have similar entropic profile. However, this assumption can be seen as an attempt to simplify the scenario, as entropy estimation of binding process still lacks direct and reliable experimental/computational methods [240]. Experimental methods seek to estimate this quantity from the conformation flexibility as proxy for it and relate NMR relaxation parameter to calibrate it with conformation part of the binding entropy; conformation entropy is again assumed to be linearly correlated with the total binding entropy [241]. While, computation methods also try to estimate configurational entropy on similar line-of-thought, using molecular fluctuation data generated from molecular mechanics as a proxy for the entropy and thereby try to estimate configurational entropy from it [242–245]. Normal mode analysis (NMA) tries to infer

conformational entropy as function of the vibration modes where DOFs are modeled as a set of simple harmonic oscillators, vibrating independently [246], but with the growing understanding of the nature of vibrational modes of biomolecules, it was realized that NMA is not the most suitable theory [247] for understanding entropy. Thus, methods utilizing internal coordinates for molecular description in conjunction with approximations representing full dimensional probability density function as a series of marginal PDFs of fluctuation of DOFs got attention of research community. This theory has been successfully applied to estimate entropy for small molecules [248], peptides [249, 250], to protein–peptide binding study with at least qualitative insight, while quantitative aspect still remains to be debatable [251, 252]. In some case, even for the set of ligands binding to the same receptor, entropic components are surprisingly quite different and play a crucial role in deciding the rank/affinity order of ligands.

As mentioned above, we found out that for a set of experimentally known ligands binding to the *P. falciparum* protein kinase PfPK5, docking scores yielded very poor correlation with experimental affinity, even inclusion of end-state free energy using MM-PBSA [253] method using 3 μ s simulation data for each of the ligands, no significant improvement in computed affinity was observed. However, when configurational entropy for the ligands was included with the MM-PBSA estimates, a significant improvement in the bonding affinity was observed (manuscript in preparation).

As shown in (Fig. 13), achieving convergence to reduce error in estimation of entropy takes longer trajectories i.e., covering larger configuration space. Using a distance cutoff-based adaptation of Maximum Information Spanning Tree (MIST) called Neighbor Approximated Maximum Information Spanning Tree (A-MIST)

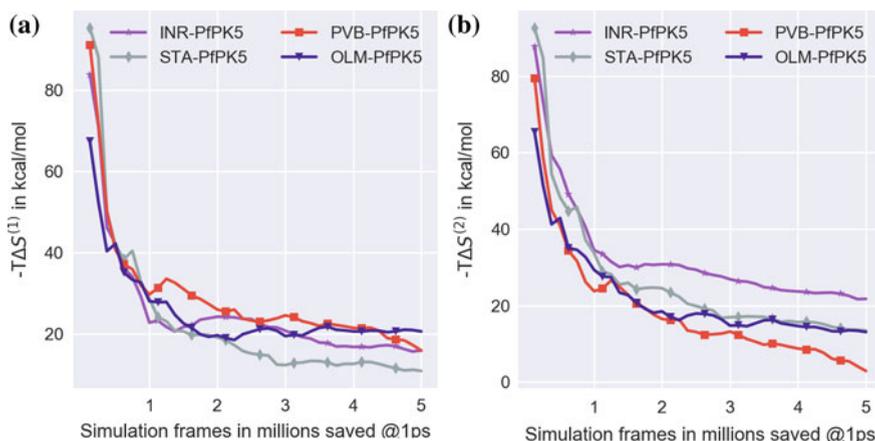


Fig. 13 Binding configurational entropy estimated using A-MIST methods with a distance cutoff of 14 Å and convergence of estimate with simulation time is shown. **a** Convergence of first order (assuming DOFs are uncorrelated) is shown. **b** Convergence of second order (accounting pair-wise correlations DOFs) is shown

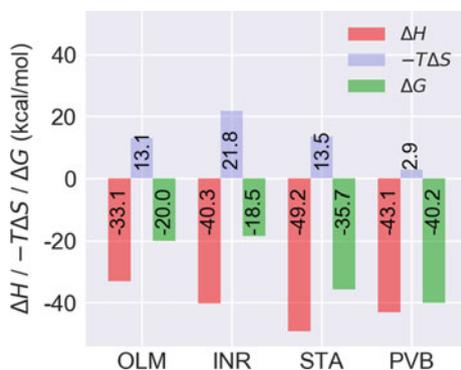


Fig. 14 Enthalpy, configurational entropy and free energy of binding of chosen inhibitors is shown in kcal/mol. Inhibitors are arranged in increasing experimental affinity ($RT\ln(\text{IC}_{50})$) order from left to right. Enthalpy is calculated using MM-PBSA method as discussed earlier. Here, temperature is taken to be 300 K

[254], the configurational entropy was estimated using MD dataset of $\sim 5 \mu\text{s}$, adding enthalpy, Free energy was calculated. It indicates that largely omitted entropic contributions can play important role and even deciding factor in case of small ligands binding to the flexible proteins (Fig. 14).

As shown in Fig. 14, combining enthalpy (ΔH) and configurational entropy ($-T\Delta S_{\text{config}}$) of binding for chosen inhibitors, the binding free energy (ΔG) for best binder PVB is highest. However, binding free energy does not discriminate between OLM and INR, where experimentally OLM is weakest binder. Lower ΔG for INR (-18.5 kcal/mol) in comparison to OLM (-20.0 kcal/mol) may be attributed to the role of solvation free energy which is not accounted rigorously in MM-PBSA methods. Variations in configurational entropy of binding from 21.8 kcal/mol to mere 2.9 kcal/mol suggest that different ligands modulate and influence receptor flexibility in their own different way while forming complex, highlighting importance of receptor flexibility in binding affinity prediction studies; recently more attentions are attracted in this field.

6.3 Thermodynamic Methods

Relative binding free energy for a ligand formed by a chemical group substitution relative to parent compound can be computed using free energy perturbation molecular dynamics simulation [255]. This technique requires constructing a path from parent ligand L_1 to analog ligand L_2 , which binds to a common receptor R, in two steps as follows. First, by carrying out a sequence of simulations in solvent and mutating L_1 to L_2 through several intermediate points and adding up the free energy changes along hypothetical intermediate points to yield free energy (say A_s) of mutating L_1 to L_2 in solvent, then similarly, mutating the ligand L_1 to L_2 in the

binding pocket of receptor in solvent to get free energy change (say A_p). Finally, subtracting A_p from A_s gives the free energy change of the binding [255]. As early as 1985, to test the concept, it was successfully applied to calculate relative solvation free energy of Cl^- and Br^- , and computed Helmholtz free energy $\Delta\Delta A$ (3.35 ± 0.15 kcal/mol) was shown to be in excellent agreement $\Delta\Delta A_{\text{hydr}} \approx \Delta\Delta G_{\text{hydr}} = 3.3$ kcal/mol with experimental value [256]. Further, the applicability of the method was extended to non-trivial systems, e.g., amino acids and their side chains, nucleic acid bases, and other small organic molecules; computed solvation free energies of these molecules are found to be in agreement with experiment [257, 258].

Relative free energy or potential of mean force (pmf, $w(r_c)$)-based methods relate it to the distribution of a chosen reaction coordinate (r_c), the direct sampling along r_c , and constructing its distribution function $g(r_c)$. The distribution function of reaction coordinate $g(r_c)$ can be related to pmf ($w(r_c)$) as

$$w(r_c) = -k_B T \ln g(r_c) + \text{constant} \quad (5)$$

However, barrier on the $w(r_c)$ can limit the sampling thereby the estimated pmf. Therefore, techniques like Umbrella sampling and Importance Sampling were introduced. But, choosing the right biasing function and ability to verify the adequacy of sampling for simulation widow is still challenging. A brief review of these methods is presented by Jorgensen et al. [259]. Statistical perturbation theory (SPT)-based methods which estimate free energy difference between systems i and j where sampling is based on system i [259]. Authors summarized several applications of SPT-based methods, e.g., for relative solvation free energy, relative pK_a values, study of solvent effect on conformational equilibria, study of binding and molecular recognition, and study of reactions in solvent [259]. The computational cost of carrying out SPT-based calculations inspired cost-effective semi-empirical methods using MD simulation samples for binding free energy calculation [260]. Aqvist et al. divided the binding free energy in two independent components electrostatic and non-polar, where electrostatic component $\Delta G_{\text{solv}}^{\text{el}}$ was taken to be half of the solvent-ion interaction energy [260]. For non-polar component, linearity between solvent size sigma and non-polar van der Waals energy and corresponding solvation energy, empirical parameter α was derived to relate vdW component of solvation free energy $\Delta G_{\text{solv}}^{\text{vdW}}$ with average of vdW component of interaction potential for transferring ligand from binding site (i) to solvent (s) given by $\Delta G_{\text{solv}}^{\text{vdW}} = \alpha \langle \Delta V_{i \rightarrow s}^{\text{vdW}} \rangle$ to yield expression for binding free energy [260] as: $\Delta G_{\text{bind}} = 1/2 \cdot \langle V_{i \rightarrow s}^{\text{el}} \rangle + \alpha \langle V_{i \rightarrow s}^{\text{vdW}} \rangle$. This new semi-empirical method was tested on aspartic protease endothiapsin and five small-molecule inhibitors with one as reference for which binding data and also crystal structure were available. It was reported that predicted relative binding free energy has mean unsigned error of 0.39 kcal/mol with highest for one of five inhibitors being 0.53 kcal/mol with parameter $\alpha = 0.161$ [260]. Application of such methods in details was discussed

by Warshel and co-workers who have systematically examined performance of protein dipoles Langevin dipoles (PDL) and other techniques using phosphorylcholine analogs binding to murine myeloma protein (McPC603) [261].

7 Molecular Recognition and Brownian Dynamics

As earlier discussed diffusional encounter of reacting substrates is the prerequisite for the binding interaction to happen [153]. Diffusional encounter is basically controlled by the long-range electrostatic interaction between participating chemical species [262]. Generally, the timescale of such encounter is from micro- to millisecond, which is tough to achieve with existing hardware technologies using molecular dynamics even for small- to moderate-sized biomolecules [263]. Therefore, simplified coarse-grained models of biomolecules can be simulated using Langevin dynamics and Brownian dynamics [262]. Brownian dynamics has been successfully applied to study ion permeations through ion channels [264] and enzymatic reactions [265]. However, to gain kinetic insight into receptor–ligand recognition, BD can be utilized [266–269], but BD being computationally very expensive is practically challenging [263]. This has called for alternate methods with simplistic approaches to study recognition process.

Supervised molecular dynamics (SuMD), a tabu-like search algorithm, aims to predict the pose of the ligand in the binding site of its cognate receptor, monitoring ligand-binding site distance along a series of short MD simulation has been proposed [225]. SuMD has been successfully applied to study a variety of molecular recognition processes [270–272]. In particular, Moro and co-workers applied to study molecular recognition process of four globular receptor–ligand systems and two transmembrane receptor ligand systems; in all these cases, experimental crystal structures and binding affinity values (IC_{50} , K_i or K_d) were already known [271]. In the study, it is observed that using SuMD, binding from unbound state (where ligand is placed at >30 – 50 Å away from binding site) of above ligands to their cognate receptor can be simulated; moreover, various interaction hot spots (metastable states) during recognition are possible to explore, which may be important in providing insight into kinetics of the recognition process, hence better designing of ligand [271]. In another study, the effect of allosteric modulator LUF6000 on adenosine binding with A_3 adenosine receptor (A_3AR) was reported. In this study, recognition of allosteric modulator LUF6000 to A_3AR and adenosine to A_3AR in presence and absence of LUF6000 was studied using SuMD. It is observed that adenosine visited a metastable site between helices EL3 and EL2, participating in hydrogen bonds with Val259 and Gln261, and it triggers an orientation change in adenosine mediated through hydrophobic interactions before occupying the binding site [270]. In future, such techniques along with Free Energy perturbation method will provide more accurate estimation of free energy binding of ligands to receptors which will include the flexibility of both partners.

8 Ligand Becomes Drug!

Drug research encompasses by various pipelines to achieve common goal, i.e., new therapeutic molecules. After the successful identification of the novel ligand or lead molecules by either computational or medicinal chemistry approach, each molecule must be characterized for absorption, distribution, metabolism, excretion, and toxicity (ADME-Tox) properties along with pharmacokinetic/pharmacodynamic (PK/PD) activity that decides the success rate of the drug [273, 274]. Evaluation of these properties belongs to the pre-clinical stage, and result of this stage decides the advancement of novel chemical entity (NCE) to clinical stage. Failure of the drug is dependent on the targeted therapeutic area; comparatively drug targeted to cardiovascular has maximum chance of success than CNS targeted [261]. So, successful candidates have to fulfill the essential criteria of potency, selectivity, oral bioavailability, therapeutic efficacy, along with an acceptable side effect profile [275]. Testing of thousands of leads molecules, found to be active against any disease, requires huge amount of money and time, and also it is not always easy to perform every test [276]. Understanding from the already prescribed drugs and knowledge from the failure rate during the different clinical stages has provided directions and specified various properties of chemicals which can be utilized to assess the lead molecules before performing costly and complex clinical tests [277].

Detailed information about ADME-Tox and its role in successful drug design is reviewed and available in many recent literatures [273, 278, 279]; however, major application of these properties is related to reduction in clinical drug failures from 40 to 10% [280]. This reduction has been seen with the advancement in the chemoinformatics and computational application in drug development process. As mentioned in the ligand design libraries, various physiochemical properties based on rules have been set to develop the lead-like and drug-like libraries to screen [281–284]. Along with these filters, for further libraries optimization filters like Pan Assay Interference Compounds (PAINS) and ALARM-NMR have been developed to remove known toxicophores or metabolically liable moieties which can interfere with the assay protocol [285, 286].

9 Summary

In this review, we have summarized many methods related to structure of receptor, characterization of active sites and subsites, binding affinity calculations, docking with specific poses, ranking chemicals and elucidated existing challenges in these methods. In spite of many mathematically and computationally elegant tools to understand and perform efficiently docking and scoring for large number of compounds, the success of identifying novel inhibitor of infectious disease and challenges thereof is still significantly high. Some of the solutions are already evident but many are yet to find. Still to ponder, how to estimate efficiently the effect of

ions, pH dependency, and Brownian dynamics, which are playing significant role in Free energy of binding to receptor. Many relevant receptors are not crystallized yet, it is clearly evident that, errors occurring in in silico model structure and plurality of interactions with the binding site play a dominant role in correctly identify any novel inhibitor. Prior knowledge of physico-chemical interactions at active site and the functional importance of interacting residues influence the pose of binding of inhibitors to flexible receptors. A prior knowledge about the mechanism of binding provides lead towards the accuracy and effective binding of docked ligand. Flexible peptides derived structures provide higher affinity and in future, emerging field of study will be designing of such restrained chemicals driven by highly active peptides. Free energy estimation, rather than scoring (however accurate it may be), provides better designing capability. Knowledge of mechanism of inhibition is mandatory for innovation of novel chemical structure to lead the drug design, even in dominant era of artificial intelligence.

In conclusion, we have attempted to highlight the existing challenges in estimating the ligand receptor binding and critically inspect the methods applied day in and day out in the field of structure-based drug design. Summarization of tools and case studies are not the scope of the review. Most important aspect is that this field evolved largely using efficient algorithm and computational tools, however, effective use requires more indulgence of chemistry and biology, in future to progress successfully.

Acknowledgements We sincerely thank the facilities under Center of Excellence and Builder project, Department of Biotechnology (DBT), for supporting the computational work and DST purse for software support. SKP is supported by DBT-BINC fellowship. The authors thank all group members and Pawan Kumar for valuable input on improving the manuscript.

References

1. Seifert MHJ, Wolf K, Vitt D (2003) Virtual high-throughput in silico screening. *Biosilico* 1:143–149. [https://doi.org/10.1016/s1478-5382\(03\)02359-x](https://doi.org/10.1016/s1478-5382(03)02359-x)
2. Engin H, Gursoy A, Nussinov R, Keskin O (2014) Network-based strategies can help mono- and poly-pharmacology drug discovery: a systems biology view. *Curr Pharm Des* 20:1201–1207. <https://doi.org/10.2174/13816128113199990066>
3. Jones PG (1984) Crystal structure determination: a critical view. *Chem Soc Rev* 13:157–172. <https://doi.org/10.1039/cs9841300157>
4. Billeter M, Wagner G, Wüthrich K (2008) Solution NMR structure determination of proteins revisited. *J Biomol NMR* 42:155–158. <https://doi.org/10.1007/s10858-008-9277-8>
5. Halperin I, Ma B, Wolfson H, Nussinov R (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins Struct Funct Genet* 47:409–443. <https://doi.org/10.1002/prot.10115>
6. Marrone TJ, Briggs JM, McCammon JA (1997) Structure-based drug design: computational advances. *Annu Rev Pharmacol Toxicol* 37:71–90. <https://doi.org/10.1146/annurev.pharmtox.37.1.71>
7. Perola E, Walters WP, Charifson PS (2004) A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins Struct Funct Genet* 56:235–249. <https://doi.org/10.1002/prot.20088>

8. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 3:935–949. <https://doi.org/10.1038/nrd1549>
9. Yuriev E, Ramsland PA (2013) Latest developments in molecular docking: 2010–2011 in review. *J Mol Recognit* 26:215–239. <https://doi.org/10.1002/jmr.2266>
10. Yuriev E, Agostino M, Ramsland PA (2011) Challenges and advances in computational docking: 2009 in review. *J Mol Recognit* 24:149–164. <https://doi.org/10.1002/jmr.1077>
11. Yuriev E, Holien J, Ramsland PA (2015) Improvements, trends, and new ideas in molecular docking: 2012–2013 in review. *J Mol Recognit* 28:581–604. <https://doi.org/10.1002/jmr.2471>
12. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein-small molecule docking methods. *J Comput Aided Mol Des* 16:151–166. <https://doi.org/10.1023/a:1020155510718>
13. Huang SY, Zou X (2010) Advances and challenges in protein-ligand docking. *Int J Mol Sci* 11:3016–3034. <https://doi.org/10.3390/ijms11083016>
14. Chen Y-C (2015) Beware of docking! *Trends Pharmacol Sci* 36:78–95. <https://doi.org/10.1016/j.tips.2014.12.001>
15. Acharya KR, Lloyd MD (2005) The advantages and limitations of protein crystal structures. *Trends Pharmacol Sci* 26:10–14. <https://doi.org/10.1016/j.tips.2004.10.011>
16. Petukh M, Stefl S, Alexov E (2013) The role of protonation states in ligand-receptor recognition and binding. *Curr Pharm Des* 19:4182–4190. <https://doi.org/10.2174/1381612811319230004>
17. Onufriev AV, Alexov E (2013) Protonation and pK changes in protein–ligand binding. *Q Rev Biophys* 46:181–209. <https://doi.org/10.1017/s0033583513000024>
18. Srivastava J, Barreiro G, Groscurth S et al (2008) Structural model and functional significance of pH-dependent talin-actin binding for focal adhesion remodeling. *Proc Natl Acad Sci U S A* 105:14436–14441. <https://doi.org/10.1073/pnas.0805163105>
19. Ferrara P, Gohlke H, Price DJ et al (2004) Assessing scoring functions for protein-ligand interactions. *J Med Chem* 47:3032–3047. <https://doi.org/10.1021/jm030489h>
20. Warren GL, Andrews CW, Capelli AM et al (2006) A critical assessment of docking programs and scoring functions. *J Med Chem* 49:5912–5931. <https://doi.org/10.1021/jm050362n>
21. Rask-Andersen M, Almén MS, Schiöth HB (2011) Trends in the exploitation of novel drug targets. *Nat Rev Drug Discov* 10:579–590. <https://doi.org/10.1038/nrd3478>
22. Finan C, Gaulton A, Kruger FA et al (2017) The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* 9:1–16. <https://doi.org/10.1126/scitranslmed.aag1166>
23. Civelek M, Lusk AJ (2014) Systems genetics approaches to understand complex traits. *Nat Rev Genet* 15:34–48. <https://doi.org/10.1038/nrg3575>
24. Csermely P, Korcsmáros T, Kiss HJM et al (2013) Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 138:333–408. <https://doi.org/10.1016/j.pharmthera.2013.01.016>
25. Zhu P, Aliabadi HM, Uludağ H, Han J (2016) Identification of potential drug targets in cancer signaling pathways using stochastic logical models. *Sci Rep* 6:23078. <https://doi.org/10.1038/srep23078>
26. Torkamani A, Topol EJ, Schork NJ (2008) Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics* 92:265–272. <https://doi.org/10.1016/j.ygeno.2008.07.011>
27. Druker BJ, Tamura S, Buchdunger E et al (1996) Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med* 2:561–566. <https://doi.org/10.1038/nm0596-561>
28. Fabian MA, Biggs WH, Treiber DK et al (2005) A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat Biotechnol* 23:329–336. <https://doi.org/10.1038/nbt1068>

29. Montgomery JA, Niwas S, Rose JD et al (1993) Structure-based design of inhibitors of purine nucleoside phosphorylase. 1. 9-(arylmethyl) derivatives of 9-deazaguanine. *J Med Chem* 36:55–69. <https://doi.org/10.1021/jm00053a008>
30. Ealick SE, Babu YS, Bugg CE et al (1993) Application of X-ray crystallographic methods in the design of purine nucleoside phosphorylase inhibitors. *Ann N Y Acad Sci* 685:237–247
31. Ealick SE, Babu YS, Bugg CE et al (1991) Application of crystallographic and modeling methods in the design of purine nucleoside phosphorylase inhibitors. *Proc Natl Acad Sci U S A* 88:11540–11544. <https://doi.org/10.1073/pnas.89.20.9974c>
32. Erion MD, Niwas S, Rose JD et al (1993) Structure-based design of inhibitors of purine nucleoside phosphorylase. 3. 9-arylmethyl derivatives of 9-deazaguanine substituted on the arylmethyl group. *J Med Chem* 36:3771–3783. <https://doi.org/10.1021/jm00076a004>
33. Freire E (2008) Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov Today* 13:869–874. <https://doi.org/10.1016/j.drudis.2008.07.005>
34. Ho M-C, Shi W, Rinaldo-Matthis A et al (2010) Four generations of transition-state analogues for human purine nucleoside phosphorylase. *Proc Natl Acad Sci* 107:4805–4812. <https://doi.org/10.1073/pnas.0913439107>
35. Edwards AA, Mason JM, Clinch K et al (2009) Altered enthalpy-entropy compensation in picomolar transition state analogues of human purine nucleoside phosphorylase. *Biochemistry* 48:5226–5238. <https://doi.org/10.1021/bi9005896>
36. Ohtaka H, Freire E (2005) Adaptive inhibitors of the HIV-1 protease. *Prog Biophys Mol Biol* 88:193–208. <https://doi.org/10.1016/j.pbiomolbio.2004.07.005>
37. Muzammil S, Armstrong AA, Kang LW et al (2007) Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations. *J Virol* 81:5144–5154. <https://doi.org/10.1128/jvi.02706-06>
38. Carbonell T, Freire E (2005) Binding thermodynamics of statins to HMG-CoA reductase. *Biochemistry* 44:11741–11748. <https://doi.org/10.1021/bi050905v>
39. Du X, Li Y, Xia Y-L et al (2016) Insights into protein-ligand interactions: mechanisms, models, and methods. *Int J Mol Sci* 17:144. <https://doi.org/10.3390/ijms17020144>
40. Amaral M., Kokh DB, Bomke J, Wegener A, Buchstaller HP, Eggenweiler HM, Matias P, Sirrenberg C, Wade RC, Frech M (2017) Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nat Commun*. <https://doi.org/10.1038/s41467-017-02258-w>
41. De Sanctis V, Kattamis C, Canatan D et al (2017) β -thalassemia distribution in the old world: an ancient disease seen from a historical standpoint. *Mediterr J Hematol Infect Dis* 9:1–14. <https://doi.org/10.4084/mjhjid.2017.018>
42. Goss C, Giardina P, Degtyaryova D et al (2014) Red blood cell transfusions for thalassemia: results of a survey assessing current practice and proposal of evidence-based guidelines. *Transfusion* 54:1773–1781. <https://doi.org/10.1111/trf.12571>
43. Michlitsch J, Walters M (2008) Recent advances in bone marrow transplantation in hemoglobinopathies. *Curr Mol Med* 8:675–689. <https://doi.org/10.2174/156652408786241393>
44. Human Hemoglobin Structures (PDB) (2018) <http://www.rcsb.org/pdb/results/results.do?tabshow=Current&qrid=205B68A9>. Accessed 15 Feb 2018
45. Brown EN, Ramaswamy S (2007) Quality of protein crystal structures. *Acta Crystallogr D Biol Crystallogr* 63:941–950. <https://doi.org/10.1107/s0907444907033847>
46. Wlodawer A, Minor W, Dauter Z, Jaskolski M (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* 275:1–21. <https://doi.org/10.1111/j.1742-4658.2007.06178.x>
47. Muller P (2009) Practical suggestions for better crystal structures. *Crystallogr Rev* 15:57–83. <https://doi.org/10.1080/08893110802547240>
48. Deller MC, Rupp B (2015) Models of protein-ligand crystal structures: trust, but verify. *J Comput Aided Mol Des* 29:817–836. <https://doi.org/10.1007/s10822-015-9833-8>
49. Carugo O (2018) How large B-factors can be in protein crystal structures. *BMC Bioinform* 19:1–9. <https://doi.org/10.1186/s12859-018-2083-8>

50. Li Z, Lazaridis T (2005) The effect of water displacement on binding thermodynamics: Concanavalin A. *J Phys Chem B* 109:662–670. <https://doi.org/10.1021/jp0477912>
51. Young T, Abel R, Kim B et al (2007) Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc Natl Acad Sci* 104:808–813. <https://doi.org/10.1073/pnas.0610202104>
52. Snyder PW, Mecinovic J, Moustakas DT et al (2011) Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *Proc Natl Acad Sci* 108:17889–17894. <https://doi.org/10.1073/pnas.1114107108>
53. Sánchez R, Sali A (1997) Advances in comparative protein-structure modelling. *Curr Opin Struct Biol* 7:206–214
54. Eswar N, John B, Mirkovic N et al (2003) Tools for comparative protein structure modeling and analysis. *Nucl Acids Res* 31:3375–3380
55. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291. <https://doi.org/10.1107/s0021889892009944>
56. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8:52–6, 29
57. McKinney JD, Zu Bentrup K Höner, Muñoz-Elias EJ et al (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* 406:735–738. <https://doi.org/10.1038/35021074>
58. Tanjore Soundarajan Balganes; Santanu Datta; Indira Ghosh (2004) WO2004087943A1
59. Nelson K, Wang FS, Boyd EF, Selander RK (1997) Size and sequence polymorphism in the isocitrate dehydrogenase kinase/phosphatase gene (*aceK*) and flanking regions in *Salmonella enterica* and *Escherichia coli*. *Genetics* 147:1509–1520
60. Garnak M, Reeves H (1979) Phosphorylation of isocitrate dehydrogenase of *Escherichia coli*. *Science* 203(80):1111–1112. <https://doi.org/10.1126/science.34215>
61. Vinekar R, Verma C, Ghosh I (2012) Functional relevance of dynamic properties of dimeric NADP-dependent isocitrate dehydrogenases. *BMC Bioinform* 13:S2. <https://doi.org/10.1186/1471-2105-13-s17-s2>
62. Vinekar R, Ghosh I (2009) Determination of phosphorylation sites for NADP-specific isocitrate dehydrogenase from *Mycobacterium tuberculosis*. *J Biomol Struct Dyn* 26:741–754. <https://doi.org/10.1080/07391102.2009.10507286>
63. Av-Gay Y, Everett M (2000) The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*. *Trends Microbiol* 8:238–244
64. Hurley JH, Thorsness PE, Ramalingam V et al (1989) Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. *Proc Natl Acad Sci U S A* 86:8635–8639. <https://doi.org/10.1073/pnas.86.22.8635>
65. Ceccarelli C, Grodsky NB, Ariyaratne N et al (2002) Crystal structure of porcine mitochondrial NADP⁺-dependent isocitrate dehydrogenase complexed with Mn²⁺ and isocitrate: Insights into the enzyme mechanism. *J Biol Chem* 277:43454–43462. <https://doi.org/10.1074/jbc.m207306200>
66. Quartararo CE, Hazra S, Hadi T, Blanchard JS (2013) Structural, kinetic and chemical mechanism of isocitrate dehydrogenase-1 from *Mycobacterium tuberculosis*. *Biochemistry* 52:1765–1775. <https://doi.org/10.1021/bi400037w>
67. Hardin C, Pogorelov TV, Luthey-Schulten Z (2002) Ab initio protein structure prediction. *Curr Opin Struct Biol* 12:176–181
68. Bonneau R, Baker D (2001) Ab initio protein structure prediction. *Annu Rev Biophys Biomol Struct* 30:173–189
69. Moulton J, Fidelis K, Kryshchukovych A et al (2017) Critical assessment of methods of protein structure prediction (CASP)—round XII. *Proteins Struct Funct Bioinform* 82:1–6
70. Singh A, Kaushik R, Mishra A et al (2016) ProTSAV: a protein tertiary structure analysis and validation server. *Biochim Biophys Acta Proteins Proteomics* 1864:11–19. <https://doi.org/10.1016/j.bbapap.2015.10.004>

71. Novoa EM, de Pouplana LR, Barril X, Orozco M (2010) Ensemble docking from homology models. *J Chem Theory Comput* 6:2547–2557
72. Kadam RU, Juraszek J, Brandenburg B et al (2017) Potent peptidic fusion inhibitors of influenza virus. *Science* 358(80):496–502. <https://doi.org/10.1126/science.aan051>
73. Chang YS, Graves B, Guerlavais V et al (2013) Stapled α -helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc Natl Acad Sci* 110:E3445–E3454. <https://doi.org/10.1073/pnas.1303002110>
74. Shadfian M, Lopez-Pajares V, Yuan Z-M (2012) MDM2 and MDMX: alone and together in regulation of p53. *Transl Cancer Res* 1:88–89. <https://doi.org/10.3978/j.issn.2218-676x.2012.04.02>
75. Tiwari G, Verma CS (2017) Toward understanding the molecular recognition of albumin by p53-activating stapled peptide ATSP-7041. *J Phys Chem B* 121:657–670. <https://doi.org/10.1021/acs.jpcc.6b09900>
76. Schindler CEM, De Vries SJ, Zacharias M (2015) Fully blind peptide-protein docking with pepATTRACT. *Structure* 23:1507–1515. <https://doi.org/10.1016/j.str.2015.05.021>
77. Leder L, Berger C, Bomhauser S et al (1995) Spectroscopic, calorimetric, and kinetic demonstration of conformational adaptation in peptide-antibody recognition. *Biochemistry* 34:16509–16518. <https://doi.org/10.1021/bi00050a035>
78. Ferrari AM, Wei BQ, Costantino L, Shoichet BK (2004) Soft docking and multiple receptor conformations in virtual screening. *J Med Chem* 47:5076–5084. <https://doi.org/10.1021/jm049756p>
79. Totrov M, Ferna J, Abagyan R (2002) Soft protein–protein docking in internal coordinates. *Protein Sci* 11:280–291. <https://doi.org/10.1110/ps.19202.ical>
80. Li CH, Ma XH, Chen WZ, Wang CX (2003) A soft docking algorithm for predicting the structure of antibody-antigen complexes. *Proteins Struct Funct Genet* 52:47–50. <https://doi.org/10.1002/prot.10382>
81. Alonso H, Bliznyuk AA, Gready JE (2006) Combining docking and molecular dynamic simulations in drug design. *Med Res Rev* 26:531–568. <https://doi.org/10.1002/med.20067>
82. Leach AR (1994) Ligand docking to proteins with discrete side-chain flexibility. *J Mol Biol* 235:345–356. [https://doi.org/10.1016/s0022-2836\(05\)80038-5](https://doi.org/10.1016/s0022-2836(05)80038-5)
83. Schnecke V, Kuhn LA (2000) Virtual screening with solvation and ligand-induced complementarity. *Perspect Drug Discov Des* 20:171–190. <https://doi.org/10.1023/a:1008737207775>
84. Källblad P, Dean PM (2003) Efficient conformational sampling of local side-chain flexibility. *J Mol Biol* 326:1651–1665. [https://doi.org/10.1016/s0022-2836\(03\)00083-4](https://doi.org/10.1016/s0022-2836(03)00083-4)
85. Frimurer TM, Peters GH, Iversen LF et al (2003) Ligand-induced conformational changes: Improved predictions of ligand binding conformations and affinities. *Biophys J* 84:2273–2281. [https://doi.org/10.1016/s0006-3495\(03\)75033-4](https://doi.org/10.1016/s0006-3495(03)75033-4)
86. Gaudreault F, Chartier M, Najmanovich R (2012) Side-chain rotamer changes upon ligand binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. *Bioinformatics* 28:423–430. <https://doi.org/10.1093/bioinformatics/bts395>
87. Apostolakis J, Plückthun A, Caffisch A (1998) Docking small ligands in flexible binding sites. *J Comput Chem* 19:21–37. [https://doi.org/10.1002/\(sici\)1096-987x\(19980115\)19:1%3c21::aid-jcc2%3e3.0.co;2-0](https://doi.org/10.1002/(sici)1096-987x(19980115)19:1%3c21::aid-jcc2%3e3.0.co;2-0)
88. Davis IW, Baker D (2009) RosettaLigand docking with full ligand and receptor flexibility. *J Mol Biol* 385:381–392. <https://doi.org/10.1016/j.jmb.2008.11.010>
89. Meiler J, Baker D (2006) ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins Struct Funct Bioinform* 65:538–548. <https://doi.org/10.1002/prot.21086>
90. Perryman AL, Lin JH, McCammon JA (2006) Optimization and computational evaluation of a series of potential active site inhibitors of the V82F/I84V drug-resistant mutant of HIV-1 protease: an application of the relaxed complex method of structure-based drug design. *Chem Biol Drug Des* 67:336–345. <https://doi.org/10.1111/j.1747-0285.2006.00382.x>

91. Shandilya A, Chacko S, Jayaram B, Ghosh I (2013) A plausible mechanism for the antimalarial activity of artemisinin: a computational approach. *Sci Rep* 3:2513
92. Li J, Zhou B (2010) Biological actions of artemisinin: insights from medicinal chemistry studies. *Molecules* 15:1378–1397
93. Eckstein-Ludwig U, Webb RJ, Van Goethem IDA et al (2003) Artemisinins target the SERCA of *Plasmodium falciparum*. *Nature* 424:957
94. Pawan K, Shandilya A, Jayaram B, Ghosh I (2016) Integrative method for finding antimalarials using in silico approach. In: Kholmurodov KT (ed) *Computer design for new drugs and materials*. Nova Science Publishers, New York, NY, pp 13–38
95. Waelbroeck M (1982) The pH dependence of insulin binding. A quantitative study. *J Biol Chem* 257:8284–8291
96. Ellis CR, Shen J (2015) pH-dependent population shift regulates BACE1 activity and inhibition. *J Am Chem Soc* 137:9543–9546. <https://doi.org/10.1021/jacs.5b05891>
97. Mongan J, Case DA, McCammon JA (2004) Constant pH molecular dynamics in generalized Born implicit solvent. *J Comput Chem* 25:2038–2048. <https://doi.org/10.1002/jcc.20139>
98. Kim MO, Blachly PG, McCammon JA (2015) Conformational dynamics and binding free energies of inhibitors of BACE-1: from the perspective of protonation equilibria. *PLoS Comput Biol* 11:1–28. <https://doi.org/10.1371/journal.pcbi.1004341>
99. Hodge CN, Aldrich PE, Bachelier LT et al (1996) Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chem Biol* 3:301–314. [https://doi.org/10.1016/s1074-5521\(96\)90110-6](https://doi.org/10.1016/s1074-5521(96)90110-6)
100. Lam PY, Jadhav PK, Eyermann CJ et al (1994) Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* 263(80):380–384. <https://doi.org/10.1126/science.8278812>
101. Kumalo HM, Bhakat S, Soliman MES (2015) Theory and applications of covalent docking in drug discovery: merits and pitfalls. *Molecules* 20:1984–2000. <https://doi.org/10.3390/molecules20021984>
102. Lecomte M, Laneuville O, Ji C et al (1994) Acetylation of human prostaglandin endoperoxide synthase-2 (cyclooxygenase-2) by aspirin. *J Biol Chem* 269:13207–13215
103. Hadváry P, Lengsfeld H, Wolfer H (1988) Inhibition of pancreatic lipase in vitro by the covalent inhibitor tetrahydropipstatin. *Biochem J* 256:357–361. <https://doi.org/10.1042/bj2560357>
104. Scarpino A, Ferenczy GG, Keserű GM (2018) Comparative evaluation of covalent docking tools. *J Chem Inf Model* acs.jcim.8b00228. <https://doi.org/10.1021/acs.jcim.8b00228>
105. Zhu K, Borrelli KW, Greenwood JR et al (2014) Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *J Chem Inf Model* 54:1932–1940. <https://doi.org/10.1021/ci500118s>
106. Bianco G, Forli S, Goodsell DS, Olson AJ (2016) Covalent docking using autodock: two-point attractor and flexible side chain methods. *Protein Sci* 25:295–301. <https://doi.org/10.1002/pro.2733>
107. Corbeil CR, Englebienne P, Moitessier N (2007) Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J Chem Inf Model* 47:435–449. <https://doi.org/10.1021/ci6002637>
108. MOE: Molecular Operating Environment (2018) http://www.chemcomp.com/MOE-Molecular_Operating_Environment.htm. Accessed 7 Feb 2018
109. Katritch V, Byrd CM, Tseitin V et al (2007) Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase I7L using homology modeling and covalent docking approaches. *J Comput Aided Mol Des* 21:549–558. <https://doi.org/10.1007/s10822-007-9138-7>
110. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Improved protein–ligand docking using GOLD. *Proteins Struct Funct Bioinform* 623:609–623. <https://doi.org/10.1002/prot.10465>

111. Domínguez JL, Christopeit T, Villaverde MC et al (2010) Effect of the protonation state of the titratable residues on the inhibitor affinity to BACE-1. *Biochemistry* 49:7255–7263. <https://doi.org/10.1021/bi100637n>
112. Peryman AL, Lin J (2004) HIV-1 protease molecular dynamics of a wild-type and of the V82F / I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci* 13:1108–1123. <https://doi.org/10.1110/ps.03468904.importance>
113. Trylska J, Tozzini V, Chang CEA, McCammon JA (2007) HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys J* 92:4179–4187. <https://doi.org/10.1529/biophysj.106.100560>
114. Liu Z-P, Wu L-Y, Wang Y et al (2008) Bridging protein local structures and protein functions. *Amino Acids* 35:627–650
115. Ratnaparkhi GS, Varadarajan R (2000) Thermodynamic and structural studies of cavity formation in proteins suggest that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. *Biochemistry* 39:12365–12374
116. DesJarlais RL, Sheridan RP, Seibel GL et al (1988) Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J Med Chem* 31:722–729
117. Anderson AC (2003) The process of structure-based drug design. *Chem Biol* 10:787–797. <https://doi.org/10.1016/j.chembiol.2003.09.002>
118. Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. *J Mol Biol* 55:379–400
119. Connolly ML (1983) Analytical molecular surface calculation. *J Appl Crystallogr* 16:548–558
120. Connolly ML (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* 221:709–713
121. Liang J, Woodward C, Edelsbrunner H (1998) Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci* 7:1884–1897
122. Edelsbrunner H (1995) Smooth surfaces for multi-scale shape representation. In: *International conference on foundations of software technology and theoretical computer science*, pp 391–412
123. Mücke EP (1998) A robust implementation for three-dimensional Delaunay triangulations. *Int J Comput Geom Appl* 8:255–276
124. Bryant R, Edelsbrunner H, Koehl P, Levitt M (2004) The area derivative of a space-filling diagram. *Discrete Comput Geom* 32:293–308
125. Dundas J, Ouyang Z, Tseng J et al (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 34:W116–W118
126. Benkaidali L, André F, Maouche B et al (2013) Computing cavities, channels, pores and pockets in proteins from non-spherical ligands models. *Bioinformatics* 30:792–800
127. Huang B, Schroeder M (2006) LIGSITE csc: predicting ligand binding sites using the Connolly surface and degree of conservation. *BMC Struct Biol* 6:19
128. Oliveira SHP, Ferraz FAN, Honorato RV et al (2014) KVFinder: steered identification of protein cavities as a PyMOL plugin. *BMC Bioinform* 15:197
129. Brady GP, Stouten PFW (2000) Fast prediction and visualization of protein binding pockets with PASS. *J Comput Aided Mol Des* 14:383–401
130. Cziriák G (2015) PrinCCes: continuity-based geometric decomposition and systematic visualization of the void repertoire of proteins. *J Mol Graph Model* 62:118–127
131. Yu J, Zhou Y, Tanaka I, Yao M (2009) Roll: a new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. *Bioinformatics* 26:46–52
132. Sheffler W, Baker D (2009) RosettaHoles: rapid assessment of protein core packing for structure prediction, refinement, design, and validation. *Protein Sci* 18:229–239
133. Laskowski RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph* 13:323–330

134. Kleywegt GJ, Jones TA (1994) Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr Sect D: Biol Crystallogr* 50:178–185
135. Jacoby E, Fauchère J-L, Raimbaud E et al (1999) A three binding site hypothesis for the interaction of ligands with monoamine α protein-coupled receptors: implications for combinatorial ligand design. *Mol Inform* 18:561–572
136. Deganutti G, Moro S (2017) Estimation of kinetic and thermodynamic ligand-binding parameters using computational strategies. *Future Med Chem* 9:507–523
137. Chiu SH, Xie L (2016) Toward high-throughput predictive modeling of protein binding/unbinding kinetics. *J Chem Inf Model* 56:1164–1174
138. Singh A (2014) Selectivity and specificity profiling of binding sites of SER/THR kinases: case study of homo sapiens and *Plasmodium falciparum*. Jawaharlal Nehru University
139. Hanks SK, Hunter T (1995) Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J* 9:576–596. <https://doi.org/10.1096/fasebj.9.8.776834>
140. Freire E (2015) The binding thermodynamics of drug candidates. In: *thermodynamics and kinetics of drug binding*. Wiley Online Library, pp 1–13
141. Cross S, Baroni M, Carosati E et al (2010) FLAP: GRID molecular interaction fields in virtual screening. Validation using the DUD data set. *J Chem Inf Model* 50:1442–1450
142. Kaalia R, Kumar A, Srinivasan A, Ghosh I (2015) An ab initio method for designing multi-target specific pharmacophores using complementary interaction field of aspartic proteases. *Mol Inform* 34:380–393
143. Nakashima R, Sakurai K, Yamasaki S et al (2013) Structural basis for the inhibition of bacterial multidrug exporters. *Nature* 500:102
144. Koshland DE (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc Natl Acad Sci* 44:98–104. <https://doi.org/10.1073/pnas.44.2.98>
145. Debler EW, Müller R, Hilvert D, Wilson IA (2008) Conformational isomerism can limit antibody catalysis. *J Biol Chem* 283:16554–16560. <https://doi.org/10.1074/jbc.m710256200>
146. Doshi U, McGowan LC, Ladani ST, Hamelberg D (2012) Resolving the complex role of enzyme conformational dynamics in catalytic function. *Proc Natl Acad Sci* 109:5699–5704. <https://doi.org/10.1073/pnas.1117060109>
147. Teague SJ (2003) Implications of protein flexibility for drug discovery. *Nat Rev Drug Discov* 2:527–541. <https://doi.org/10.1038/nrd1129>
148. Hawkins PCD (2017) Conformation generation: the state of the art. *J Chem Inf Model* 57:1747–1756. <https://doi.org/10.1021/acs.jcim.7b00221>
149. Grinter SZ, Zou X (2014) Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design. *Molecules* 19:10150–10176. <https://doi.org/10.3390/molecules190710150>
150. Mobley DL, Dill KA (2009) Binding of small-molecule ligands to proteins: “what you see” is not always “what you get.” *Structure* 17:489–498
151. London N, Movshovitz-Attias D, Schueler-Furman O (2010) The structural basis of peptide-protein binding strategies. *Structure* 18:188–199. <https://doi.org/10.1016/j.str.2009.11.012>
152. London N, Raveh B, Schueler-Furman O (2013) Peptide docking and structure-based characterization of peptide binding: from knowledge to know-how. *Curr Opin Struct Biol* 23:894–902. <https://doi.org/10.1016/j.sbi.2013.07.006>
153. Bernetti M, Cavalli A, Mollica L (2017) Protein–ligand (un)binding kinetics as a new paradigm for drug discovery at the crossroad between experiments and modelling. *Med Chem Commun* 8:534–550. <https://doi.org/10.1039/c6md00581k>
154. Pan AC, Borhani DW, Dror RO, Shaw DE (2013) Molecular determinants of drug-receptor binding kinetics. *Drug Discov Today* 18:667–673. <https://doi.org/10.1016/j.drudis.2013.02.007>
155. Case DA (1988) Dynamical simulation of rate constants in protein-ligand interactions. *Prog Biophys Mol Biol* 52:39–70. [https://doi.org/10.1016/0079-6107\(88\)90007-7](https://doi.org/10.1016/0079-6107(88)90007-7)

156. Liu J, Wang R (2015) Classification of current scoring functions. *J Chem Inf Model* 55:475–482. <https://doi.org/10.1021/ci500731a>
157. Cornell WD, Cieplak P, Bayly CI et al (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
158. MacKerell AD Jr, Bashford D, Bellott M et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
159. Allen WJ, Balias TE, Mukherjee S et al (2015) DOCK 6: impact of new features and current docking performance. *J Comput Chem* 36:1132–1156. <https://doi.org/10.1002/jcc.23905>
160. Simonson T, Archontis G, Karplus M (2002) Free energy simulations come of age: protein–ligand recognition. *Acc Chem Res* 35:430–437
161. Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FLEXX incremental construction algorithm for protein–ligand docking. *Proteins Struct Funct Bioinform* 37:228–241
162. Böhm H-J (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* 8:243–256
163. Eldridge MD, Murray CW, Auton TR et al (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J Comput Aided Mol Des* 11:425–445
164. Böhm H-J (1992) LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J Comput Aided Mol Des* 6:593–606
165. Friesner RA, Banks JL, Murphy RB et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* 47:1739–1749. <https://doi.org/10.1021/jm0306430>
166. Wang R, Lai L, Wang S (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J Comput Aided Mol Des* 16:11–26
167. Sippl MJ (1995) Knowledge-based potentials for proteins. *Curr Opin Struct Biol* 5:229–235
168. Lyne PD (2002) Structure-based virtual screening: an overview. *Drug Discov Today* 7:1047–1055. [https://doi.org/10.1016/s1359-6446\(02\)02483-2](https://doi.org/10.1016/s1359-6446(02)02483-2)
169. Muegge I (2000) A knowledge-based scoring function for protein–ligand interactions: probing the reference state. *Perspect Drug Discov Des* 20:99–114
170. Gohlke H, Hendlich M, Klebe G (2000) Knowledge-based scoring function to predict protein–ligand interactions. *J Mol Biol* 295:337–356
171. Head RD, Smythe ML, Oprea TI et al (1996) VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J Am Chem Soc* 118:3959–3969
172. Rarey M, Kramer B, Lengauer T, Klebe G (1996) A fast flexible docking method using an incremental construction algorithm. *J Mol Biol* 261:470–489
173. Morris GM, Goodsell DS, Halliday RS et al (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J Comput Chem* 19:1639–1662
174. Spyraakis F, Amadasi A, Fornabaio M et al (2007) The consequences of scoring docked ligand conformations using free energy correlations. *Eur J Med Chem* 42:921–933. <https://doi.org/10.1016/j.ejmech.2006.12.037>
175. Oprea TI, Marshall GR (1998) Receptor-based prediction of binding affinities. *Perspect Drug Discov Des* 9:35–61
176. Huang N, Shoichet BK, Irwin JJ (2006) Benchmarking sets for molecular docking. *J Med Chem* 49:6789–6801. <https://doi.org/10.1021/jm0608356>
177. Empereur-Mot C, Guillemain H, Latouche A et al (2015) Predictiveness curves in virtual screening. *J Cheminform* 7:52. <https://doi.org/10.1186/s13321-015-0100-8>
178. Bauer MR, Ibrahim TM, Vogel SM, Boeckler FM (2013) Evaluation and optimization of virtual screening workflows with DEKOIS 2.0—a public library of challenging docking benchmark sets. *J Chem Inf Model* 53:1447–1462

179. Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. *Nucl Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
180. Gore S, Sanz García E, Hendrickx PMS et al (2017) Validation of structures in the protein data bank. *Structure* 25:1916–1927. <https://doi.org/10.1016/j.str.2017.10.009>
181. Morris GM, Huey R, Lindstrom W et al (2009) AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 30:2785–2791. <https://doi.org/10.1002/jcc.21256>
182. Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J Mol Biol* 428:720–725. <https://doi.org/10.1016/j.jmb.2015.09.014>
183. Diller DJ, Merz KM (2001) High throughput docking for library design and library prioritization. *Proteins Struct Funct Genet* 43:113–124. [https://doi.org/10.1002/1097-0134\(20010501\)43:2%3c113:aid-prot1023%3e3.0.co;2-t](https://doi.org/10.1002/1097-0134(20010501)43:2%3c113:aid-prot1023%3e3.0.co;2-t)
184. Venkatachalam CM, Jiang X, Oldfield T, Waldman M (2003) LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. *J Mol Graph Model* 21:289–307. [https://doi.org/10.1016/s1093-3263\(02\)00164-x](https://doi.org/10.1016/s1093-3263(02)00164-x)
185. Holton S, Merckx A, Burgess D et al (2003) Structures of *P. falciparum* PfPK5 test the CDK regulation paradigm and suggest mechanisms of small molecule inhibition. *Structure* 11:1329–1337. <https://doi.org/10.1016/j.str.2003.09.020>
186. Engels MFM, Gibbs AC, Jaeger EP et al (2006) A cluster-based strategy for assessing the overlap between large chemical libraries and its application to a recent acquisition. *J Chem Inf Model* 46:2651–2660. <https://doi.org/10.1021/ci600219n>
187. Krier M, Bret G, Rognan D (2006) Assessing the scaffold diversity of screening libraries. *J Chem Inf Model* 46:512–524. <https://doi.org/10.1021/ci050352v>
188. McGregor MJ, Muskal SM (1999) Pharmacophore fingerprinting. 1. Application to QSAR and focused library design. *J Chem Inf Comput Sci* 39:569–574. <https://doi.org/10.1021/ci980159j>
189. Reymond J-L, van Deursen R, Blum LC, Ruddigkeit L (2010) Chemical space as a source for new drugs. *Medchemcomm* 1:30. <https://doi.org/10.1039/c0md00020e>
190. Williams AJ (2008) A perspective of publicly accessible/open-access chemistry databases. *Drug Discov Today* 13:495–501. <https://doi.org/10.1016/j.drudis.2008.03.017>
191. Lavecchia A, Di Giovanni C (2013) Virtual screening strategies in drug discovery: a critical review. *Curr Med Chem* 20:2839–2860. <https://doi.org/10.2174/09298673113209990001>
192. Hann MM, Oprea TI (2004) Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol* 8:255–263. <https://doi.org/10.1016/j.cbpa.2004.04.003>
193. Villoutreix Bruno O, Renault Nicolas, Lagorce David et al (2007) Free resources to assist structure-based virtual ligand screening experiments. *Curr Protein Pept Sci* 8:381–411. <https://doi.org/10.2174/138920307781369391>
194. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2012) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 64:4–17. <https://doi.org/10.1016/j.addr.2012.09.019>
195. Congreve M, Carr R, Murray C, Jhoti H (2003) A “rule of three” for fragment-based lead discovery? *Drug Discov Today* 8:876–877. [https://doi.org/10.1016/s1359-6446\(03\)02831-9](https://doi.org/10.1016/s1359-6446(03)02831-9)
196. Hughes JD, Blagg J, Price DA et al (2008) Physicochemical drug properties associated with in vivo toxicological outcomes. *Bioorg Med Chem Lett* 18:4872–4875. <https://doi.org/10.1016/j.bmcl.2008.07.071>
197. Lagarde N, Zagury J-F, Montes M (2015) Benchmarking data sets for the evaluation of virtual ligand screening methods: review and perspectives. *J Chem Inf Model* 55:1297–1307. <https://doi.org/10.1021/acs.jcim.5b00090>
198. Réau M, Langenfeld F, Zagury J-F et al (2018) Decoys selection in benchmarking datasets: overview and perspectives. *Front Pharmacol* 9 <https://doi.org/10.3389/fphar.2018.00011>
199. Clark RD, Strizhev A, Leonard JM et al (2002) Consensus scoring for ligand/protein interactions. *J Mol Graph Model* 20:281–295

200. Cox JAG, Mugumbate G, Del Peral LVG et al (2016) Novel inhibitors of *Mycobacterium tuberculosis* GuaB2 identified by a target based high-throughput phenotypic screen. *Sci Rep* 6:1–10. <https://doi.org/10.1038/srep38986>
201. McInnes C (2007) Virtual screening strategies in drug discovery. *Curr Opin Chem Biol* 11:494–502. <https://doi.org/10.1016/j.cbpa.2007.08.033>
202. Kumar A, Zhang KYJ (2018) A cross docking pipeline for improving pose prediction and virtual screening performance. *J Comput Aided Mol Des* 32:163–173. <https://doi.org/10.1007/s10822-017-0048-z>
203. Thilagavathi R, Mancera RL (2010) Ligand- protein cross-docking with water molecules. *J Chem Inf Model* 50:415–421
204. Kilambi KP, Gray JJ (2017) Structure-based cross-docking analysis of antibody–antigen interactions. *Sci Rep* 7:8145
205. Kroemer RT, Vulpetti A, McDonald JJ et al (2004) Assessment of docking poses: interactions-based accuracy classification (IBAC) versus crystal structure deviations. *J Chem Inf Comput Sci* 44:871–881
206. Smith RD, Dunbar JB, Ung PM et al (2011) CSAR benchmark exercise of 2010: combined evaluation across all submitted scoring functions. *J Chem Inf Model* 51:2115–2131. <https://doi.org/10.1021/ci200269q>
207. Oda A, Tsuchida K, Takakura T et al (2006) Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J Chem Inf Model* 46:380–391. <https://doi.org/10.1021/ci050283k>
208. Dixon SL, Smondyrev AM, Knoll EH et al (2006) PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des* 20:647–671. <https://doi.org/10.1007/s10822-006-9087-6>
209. Verma J, Khedkar VM, Coutinho EC (2010) 3D-QSAR in drug design—a review. *Curr Top Med Chem* 10:95–115. <https://doi.org/10.2174/156802610790232260>
210. Halgren TA, Murphy RB, Friesner RA et al (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem* 47:1750–1759. <https://doi.org/10.1021/jm030644s>
211. Hawkins PCD, Skillman AG, Nicholls A (2007) Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* 50:74–82. <https://doi.org/10.1021/jm0603365>
212. Hawkins PCD, Skillman AG, Warren GL et al (2010) Conformer generation with OMEGA: algorithm and validation using high quality structures from the protein databank and Cambridge structural database. *J Chem Inf Model* 50:572–584. <https://doi.org/10.1021/ci100031x>
213. Takagi T, Amano M, Tomimoto M (2009) Novel method for the evaluation of 3D conformation generators. *J Chem Inf Model* 49:1377–1388. <https://doi.org/10.1021/ci800393w>
214. Ebejer J-P, Morris GM, Deane CM (2012) Freely available conformer generation methods: how good are they? *J Chem Inf Model* 52:1146–1158. <https://doi.org/10.1021/ci2004658>
215. Loferer MJ, Kolossváry I, Aszódi A (2007) Analyzing the performance of conformational search programs on compound databases. *J Mol Graph Model* 25:700–710. <https://doi.org/10.1016/j.jm gm.2006.05.008>
216. Vainio MJ, Johnson MS (2007) Generating conformer ensembles using a multiobjective genetic algorithm. *J Chem Inf Model* 47:2462–2474. <https://doi.org/10.1021/ci6005646>
217. Li J, Ehlers T, Sutter J et al (2007) CAESAR: a new conformer generation algorithm based on recursive buildup and local rotational symmetry consideration. *J Chem Inf Model* 47:1923–1932. <https://doi.org/10.1021/ci700136x>
218. Watts KS, Dalal P, Murphy RB et al (2010) ConfGen: a conformational search method for efficient generation of bioactive conformers. *J Chem Inf Model* 50:534–546. <https://doi.org/10.1021/ci100015j>

219. O'Boyle N, Vandermeersch T, Hutchison G (2011) Confab—generation of diverse low energy conformers. *J Cheminform* 3:P32. <https://doi.org/10.1186/1758-2946-3-s1-p32>
220. Gasteiger J, Rudolph C, Sadowski J (1990) Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput Methodol* 3:537–547. [https://doi.org/10.1016/0898-5529\(90\)90156-3](https://doi.org/10.1016/0898-5529(90)90156-3)
221. Riniker S, Landrum GA (2015) Better informed distance geometry: using what we know to improve conformation generation. *J Chem Inf Model* 55:2562–2574. <https://doi.org/10.1021/acs.jcim.5b00654>
222. Miteva MA, Guyon F, Tuffery P (2010) Frog2: efficient 3D conformation ensemble generator for small compounds. *Nucl Acids Res* 38:W622–W627. <https://doi.org/10.1093/nar/gkq325>
223. Sauton N, Lagorce D, Villoutreix BO, Miteva MA (2008) MS-DOCK: accurate multiple conformation generator and rigid docking protocol for multi-step virtual ligand screening. *BMC Bioinform* 9:1–12. <https://doi.org/10.1186/1471-2105-9-184>
224. Greg L (2018) RDKit: open-source cheminformatics. <http://www.rdkit.org/>. Accessed 7 Feb 2018
225. Sabbadin D, Moro S (2014) Supervised molecular dynamics (SuMD) as a helpful tool to depict GPCR-ligand recognition pathway in a nanosecond time scale. *J Chem Inf Model* 54:372–376. <https://doi.org/10.1021/ci400766b>
226. Mobley DL, Gilson MK (2017) Predicting binding free energies: frontiers and benchmarks. *Annu Rev Biophys* 46:531–558. <https://doi.org/10.1146/annurev-biophys-070816-033654>
227. Chodera JD, Mobley DL, Shirts MR et al (2011) Alchemical free energy methods for drug discovery: progress and challenges. *Curr Opin Struct Biol* 21:150–160. <https://doi.org/10.1016/j.sbi.2011.01.011>
228. Shirts MR, Mobley DL, Brown SP (2010) Free-energy calculations in structure-based drug design. In: Merz KM, Ringe D, Reynolds CH (eds) *Drug design: structure-and ligand-based approaches*. Cambridge University Press, New York, Cambridge, pp 61–86
229. Hansen N, Van Gunsteren WF (2014) Practical aspects of free-energy calculations: a review. *J Chem Theory Comput* 10:2632–2647. <https://doi.org/10.1021/ct500161f>
230. Christ CD, Mark AE, van Gunsteren WF (2009) Basic ingredients of free energy calculations: a review. *J Comput Chem* 31:1569–1582. <https://doi.org/10.1002/jcc.21450>
231. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 10:449–461. <https://doi.org/10.1517/17460441.2015.1032936>
232. Homeyer N, Gohlke H (2012) Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. *Mol Inform* 31:114–122. <https://doi.org/10.1002/minf.201100135>
233. Case DA, Babin V, Berryman JT, Betz RM, Cai Q, Cerutti DS, Cheatham TE III, Darden TA, Duke RE, Gohlke H, Goetz AW, Gusarov S, Homeyer N, Janowski P, Kaus J, Kolossváry I, Kovalenko A, Lee TS, LeGrand S, Luchko T, Luo R, Madej B, Merz KM, Paesani F, Roe DR, Roitberg A, Sagui C, Salomon-Ferrer R, Sebabra G, Simmerling CL, Smith W, Swails J, Walker RC, Wang J, Wolf RM, Wolf X, Kollman PA (2014) *Amber14 Reference manual*
234. Sprenger KG, Jaeger VW, Pfaendtner J (2015) The general AMBER force field (GAFF) can accurately predict thermodynamic and transport properties of many ionic liquids. *J Phys Chem B* 119:5882–5895. <https://doi.org/10.1021/acs.jpcc.5b00689>
235. Maier JA, Martinez C, Kasavajhala K et al (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* 11:3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255>
236. Nguyen CN, Kurtzman Young T, Gilson MK (2012) Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *J Chem Phys* 137:044101. <https://doi.org/10.1063/1.4733951>
237. Kräutler V, Van Gunsteren WF, Hünenberger PH (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations.

- J Comput Chem 22:501–508. [https://doi.org/10.1002/1096-987x\(20010415\)22:5%3c501:aid-jcc1021%3e3.0.co;2-v](https://doi.org/10.1002/1096-987x(20010415)22:5%3c501:aid-jcc1021%3e3.0.co;2-v)
238. Salomon-Ferrer R, Case DA, Walker RC (2013) An overview of the amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci* 3:198–210. <https://doi.org/10.1002/wcms.1121>
239. Miller BR, McGee TD, Swails JM et al (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput* 8:3314–3321. <https://doi.org/10.1021/ct300418h>
240. Marlow MS, Dogan J, Frederick KK et al (2010) The role of conformational entropy in molecular recognition by calmodulin. *Nat Chem Biol* 6:352–358. <https://doi.org/10.1038/nchembio.347>
241. Kasinath V, Sharp KA, Wand AJ (2013) Microscopic insights into the NMR relaxation-based protein conformational entropy meter. *J Am Chem Soc* 135:15092–15100. <https://doi.org/10.1021/ja405200u>
242. Diehl C, Engström O, Delaine T et al (2010) Protein flexibility and conformational entropy in ligand design targeting the carbohydrate recognition domain of galectin-3. *J Am Chem Soc* 132:14577–14589. <https://doi.org/10.1021/ja105852y>
243. Fenley AT, Muddana HS, Gilson MK (2012) Entropy-enthalpy transduction caused by conformational shifts can obscure the forces driving protein-ligand binding. *Proc Natl Acad Sci U S A* 109:20006–20011. <https://doi.org/10.1073/pnas.1213180109>
244. Chodera JD, Mobley DL (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu Rev Biophys* 42:121–142. <https://doi.org/10.1146/annurev-biophys-083012-130318>
245. Olsson TSG, Ladbury JE, Pitt WR, Williams MA (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. *Protein Sci* 20:1607–1618. <https://doi.org/10.1002/pro.692>
246. López-Blanco JR, Miyashita O, Tama F, Chacón P (2014) Normal mode analysis techniques in structural biology. In: John Wiley & Sons Ltd (ed) eLS. John Wiley & Sons, Ltd, Chichester, UK, p 9
247. Numata J, Wan M, Knapp E-W (2007) Conformational entropy of biomolecules: beyond the quasi-harmonic approximation. *Genome Inform* 18:192–205
248. Killian BJ, Yundenfreund Kravitz J, Gilson MK (2007) Extraction of configurational entropy from molecular simulations via an expansion approximation. *J Chem Phys* 127:024107. <https://doi.org/10.1063/1.2746329>
249. Numata J, Knapp E-W (2012) Balanced and bias-corrected computation of conformational entropy differences for molecular trajectories. *J Chem Theory Comput* 8:1235–1245. <https://doi.org/10.1021/ct200910z>
250. Suárez E, Díaz N, Méndez J, Suárez D (2013) CENCALC: a computational tool for conformational entropy calculations from molecular simulations. *J Comput Chem* 34:2041–2054. <https://doi.org/10.1002/jcc.23350>
251. Killian BJ, Kravitz JY, Somani S et al (2009) Configurational entropy in protein-peptide binding: computational study of Tsg101 ubiquitin E2 variant domain with an HIV-derived PTAP nonapeptide. *J Mol Biol* 389:315–335. <https://doi.org/10.1016/j.jmb.2009.04.003>
252. Fenley AT, Killian BJ, Hnizdo V et al (2014) Correlation as a determinant of configurational entropy in supramolecular and protein systems. *J Phys Chem B* 118:6447–6455. <https://doi.org/10.1021/jp411588b>
253. Fogolari F, Brigo a, Molinari H (2002) The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. *J Mol Recognit* 15:377–392. <https://doi.org/10.1002/jmr.577>
254. King BM, Silver NW, Tidor B (2012) Efficient calculation of molecular configurational entropies using an information theoretic approximation. *J Phys Chem B* 116:2891–2904. <https://doi.org/10.1021/jp2068123>
255. Tembre BL, Mc Cammon JA (1984) Ligand-receptor interactions. *Comput Chem* 8:281–283. [https://doi.org/10.1016/0097-8485\(84\)85020-2](https://doi.org/10.1016/0097-8485(84)85020-2)

256. Lybrand TP, Ghosh I, McCammon JA (1985) Hydration of chloride and bromide anions: determination of relative free energy by computer simulation. *J Am Chem Soc* 107:7793–7794. <https://doi.org/10.1021/ja00311a112>
257. Bash P, Singh U, Langridge R, Kollman P (1987) Free energy calculations by computer simulation. *Science* 236(80):564–568. <https://doi.org/10.1126/science.3576184>
258. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. *Chem Rev* 93:2395–2417. <https://doi.org/10.1021/cr00023a004>
259. Jorgensen WL (1989) Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc Chem Res* 22:184–189. <https://doi.org/10.1021/ar00161a004>
260. Aqvist J, Medina C, Samuelsson JE (1994) A new method for predicting binding affinity in computer-aided drug design. *Protein Eng* 7:385–391
261. Lee FS, Chu ZT, Bolger MB, Warshel A (1992) Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Eng* 5:215–228. <https://doi.org/10.1093/protein/5.3.215>
262. Ermak DL, McCammon JA (1978) Brownian dynamics with hydrodynamic interactions. *J Chem Phys* 69:1352. <https://doi.org/10.1063/1.436761>
263. Adcock SA, McCammon JA (2006) Molecular dynamics: survey of methods for simulating the activity of proteins. *Chem Rev* 106:1589–1615. <https://doi.org/10.1021/cr040426m>
264. Bek S, Jakobsson E (1994) Brownian dynamics study of a multiply-occupied cation channel: application to understanding permeation in potassium channels. *Biophys J* 66:1028–1038. [https://doi.org/10.1016/s0006-3495\(94\)80884-7](https://doi.org/10.1016/s0006-3495(94)80884-7)
265. Sines J, Allison S, McCammon JA (1990) Brownian dynamics simulation of the superoxide-superoxide dismutase reaction: iron and manganese enzymes. *J Phys Chem* 94:959–961
266. Northrup SH, Erickson HP (1992) Kinetics of protein–protein association explained by Brownian dynamics computer simulation. *Proc Natl Acad Sci U S A* 89:3338–3342
267. Kozack RE, Subramaniam S (1993) Brownian dynamics simulations of molecular recognition in an antibody-antigen system. *Protein Sci* 2:915–926. <https://doi.org/10.1002/pro.5560020605>
268. Gabdouliline RR, Wade RC (1997) Simulation of the diffusional association of barnase and barstar. *Biophys J* 72:1917–1929. [https://doi.org/10.1016/s0006-3495\(97\)78838-6](https://doi.org/10.1016/s0006-3495(97)78838-6)
269. Gabdouliline RR, Wade RC (1998) Brownian dynamics simulation of protein–protein diffusional encounter. *Methods* 14:329–341. <https://doi.org/10.1006/meth.1998.0588>
270. Deganutti G, Cuzzolin A, Ciancetta A, Moro S (2015) Understanding allosteric interactions in G protein-coupled receptors using supervised molecular dynamics: a prototype study analysing the human A3 adenosine receptor positive allosteric modulator LUF6000. *Bioorg Med Chem* 23:4065–4071. <https://doi.org/10.1016/j.bmc.2015.03.039>
271. Cuzzolin A, Sturlese M, Deganutti G et al (2016) Deciphering the complexity of ligand-protein recognition pathways using supervised molecular dynamics (SuMD) simulations. *J Chem Inf Model* 56:687–705. <https://doi.org/10.1021/acs.jcim.5b00702>
272. Paoletta S, Sabbadin D, von Kügelgen I et al (2015) Modeling ligand recognition at the P2Y12 receptor in light of X-ray structural information. *J Comput Aided Mol Des* 29:737–756. <https://doi.org/10.1007/s10822-015-9858-z>
273. Lin JH, Lu AYH (1997) Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol Rev* 49:403–449
274. Gallo JM (2010) Pharmacokinetic/pharmacodynamic-driven drug development. *Mt Sinai J Med A J Transl Pers Med* 77:381–388. <https://doi.org/10.1002/msj.20193>
275. Alavijeh MS, Chishty M, Qaiser MZ, Palmer AM (2005) Drug metabolism and pharmacokinetics, the blood-brain barrier, and central nervous system drug discovery. *NeuroRx* 2:554–571. <https://doi.org/10.1602/neurorx.2.4.554>
276. Altshuler J, Flanagan A, Guy P et al (2001) A revolution in R&D: how genomics and genetics are transforming the biopharmaceutical industry. Boston Consulting Group, Boston

277. Davis AM, Riley RJ (2004) Predictive ADMET studies, the challenges and the opportunities. *Curr Opin Chem Biol* 8:378–386
278. White RE (1998) Short-and long-term projections about the use of drug metabolism in drug discovery and development. *Drug Metab Dispos* 26:1213–1216
279. Eddershaw PJ, Beresford AP, Bayliss MK (2000) ADME/PK as part of a rational approach to drug discovery. *Drug Discov Today* 5:409–414
280. Kola I, Landis J (2004) Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 3:711
281. Harris CJ, Hill RD, Sheppard DW et al (2011) The design and application of target-focused compound libraries. *Comb Chem High Throughput Screen* 14:521–531
282. Paricharak S, Méndez-Lucio O, Chavan Ravindranath A et al (2016) Data-driven approaches used for compound library design, hit triage and bioactivity modeling in high-throughput screening. *Brief Bioinform* 19(2):277–285. <https://doi.org/10.1093/bib/bbw105>
283. Chuprina A, Lukin O, Demoiseaux R et al (2010) Drug-and lead-likeness, target class, and molecular diversity analysis of 7.9 million commercially available organic compounds provided by 29 suppliers. *J Chem Inf Model* 50:470–479
284. Oprea TI, Allu TK, Fara DC et al (2007) Lead-like, drug-like or “Pub-like”: how different are they? *J Comput Aided Mol Des* 21:113–119
285. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J Med Chem* 53:2719–2740. <https://doi.org/10.1021/jm901137j>
286. Metz JT, Huth JR, Hajduk PJ (2007) Enhancement of chemical rules for predicting compound reactivity towards protein thiol groups. *J Comput Aided Mol Des* 21:139–144. <https://doi.org/10.1007/s10822-007-9109-z>